

Langzeitarchivierung von Dokumenten aus dem World Wide Web

Diplomarbeit

im

Studiengang Bibliotheks- und Medienmanagement Diplom
Fachhochschule Stuttgart – Hochschule der Medien

Marion Mielke

Erstprüferin:

Prof. Margarete Payer

Zweitprüfer:

Prof. Bernward Hoffmann

Bearbeitungszeitraum: 08.07.2004 bis 08.10.2004

Stuttgart, Oktober 2004

Kurzfassung

Die Veröffentlichung digitaler Dokumente, die über das World Wide Web (WWW) zur Verfügung stehen, nimmt stetig zu. Hieraus ergibt sich für Bibliotheken eine neue Herausforderung: die Langzeitarchivierung und die Langzeitverfügbarkeit dieser Dokumente sicherzustellen. Insbesondere die Nationalbibliothek und die Regionalbibliotheken haben einen Sammelauftrag zu erfüllen, der in Zukunft auch auf entsprechende Netzpublikationen ausgedehnt werden soll. Hierzu ist es erforderlich technische und organisatorische Konzepte zu erarbeiten und zu erproben. In dieser Diplomarbeit werden die Lösungsmöglichkeiten, die bis jetzt gefunden wurden, und die Probleme, die diese Thematik mit sich bringt, beschrieben.

Schlagwörter

Langzeitarchivierung, Langzeitverfügbarkeit, digitale Dokumente,
Netzpublikationen, Bibliothek

Abstract

The publication of digital documents which are available about the World Wide Web (WWW) steadily increases. A new challenge surrenders for libraries: the long-term preservation and the long-term availability of these documents. In particular the national library and the regional libraries have to fulfil a collective order which should be expanded in future also to certain net-publications. Moreover it is necessary to create technical and organizational drafts and to test them. This diploma thesis shows the possibilities of solutions which were found till present, and describes the problems which this topic implicates.

Keywords

long-term preservation, long-term availability, digital documents, net-publications,
library

Inhaltsverzeichnis

Kurzfassung	2
Schlagwörter.....	2
Abstract	3
Keywords	3
Inhaltsverzeichnis	4
Abbildungsverzeichnis	6
Tabellenverzeichnis	7
1 Einleitung	8
2 Der Auftrag	9
3 Digitale Dokumente	11
3.1 Gängige Dateiformate und ihre Besonderheiten	14
3.1.1 Text-Dokumente	14
3.1.1.1 PS (PostScript)	14
3.1.1.2 PDF (Portable Document Format)	14
3.1.1.3 DOC (Word-Dokument)	15
3.1.1.4 LaTeX	15
3.1.1.5 SGML (Standard Generalized Markup Language)	15
3.1.1.6 XML (Extensible Markup Language).....	16
3.1.1.7 HTML (Hyper Text Markup Language)	16
3.1.2 Nicht-Text-Dokumente	16
3.1.2.1 JPEG (Joint Photographic Experts Group).....	16
3.1.2.2 GIF (Graphics Interchange Format).....	16
3.1.2.3 PNG (Portable Network Graphics).....	17
3.1.2.4 MP3 (MPEG-1/2 Audio Layer-3).....	17
3.2 Dokumentenstandards.....	18
3.3 Vor- und Nachteile von digitalen Dokumenten	19

4	Organisatorische Konzepte	22
4.1	Das OAIS-Referenzmodell (OAIS-RM)	24
4.1.1	Das Umgebungsmodell	24
4.1.2	Das Informationsmodell	25
4.1.3	Das Prozessmodell	29
4.2	Das DSEP-Prozessmodell	32
5	Weitere organisatorische Aspekte	38
5.1	Erschließung von Netzpublikationen	38
5.2	Metadaten	40
5.3	Authentizität und Sicherheit	44
5.4	Persistent Identifiers – URN	45
6	Strategien der Langzeitarchivierung	49
6.1	Das "Computer Museum"	50
6.2	Migration	50
6.3	Emulation	52
7	Rechtliche Aspekte	54
8	Projekte	57
8.1	NESTOR	57
8.2	Dissertationen Online	60
8.3	BOA (Baden-Württembergisches Online-Archiv)	65
8.4	SpringerLink	67
9	Schlussbetrachtung	69
	Literaturverzeichnis	70
	Erklärung	82

Abbildungsverzeichnis

Abbildung 1: OAIS-Umgebungsmodell (CCSDS 2002b, S. 2).....	25
Abbildung 2: OAIS-Informationsmodell (Borghoff u. a. 2003, S. 27)	26
Abbildung 3: Information Package (CCSDS 2002b, S. 3)	26
Abbildung 4: OAIS-Informationsmodell (OCLC/RLG 2001, S. 11)	28
Abbildung 5: OAIS-Prozessmodell (CCSDS 2002b, S. 6)	29
Abbildung 6: DSEP-Prozessmodell (van der Werf 2000, S. 21).....	33
Abbildung 7: Workflow des DSEP (Liegmann 2001)	35
Abbildung 8: Nestor (Kompetenznetzwerk Langzeitarchivierung 2004)	59
Abbildung 9: Statistik Online-Hochschulschriften (DDB 2004f)	63

Tabellenverzeichnis

Tabelle 1: Jährliches Wachstum von Websites (OCLC 2002)	20
Tabelle 2: Jährliches Wachstum der URNs (EPICUR 2004b)	49

1 Einleitung

Mit der stetig steigenden Anzahl von digitalen Dokumenten stellt sich zunehmend die Frage wie dieses kulturelle Erbe erhalten werden kann. Deshalb ist der Aufbau eines Konzepts zu Langzeitarchivierung von digitalen Dokumenten gefordert. Unter Langzeitarchivierung wird die Langzeitverfügbarkeit digitaler Ressourcen verstanden. Diese ist auf keinen bestimmten Zeitraum festgelegt. Diese Problematik ist ein hochaktuelles Thema und es werden nun schnellstmöglichst Lösungen gesucht, da es sonst zu einem Informationsverlust kommt, wie es z. B. 2001 bei der Präsidentschaftsübernahme durch Bush schon geschehen ist¹. Insbesondere betrifft dies Bibliotheken, Archive und Museen, da diese einen Sammelauftrag haben. Die dauerhafte Archivierung von digitalen Dokumenten ist zur Zeit noch mit verschiedenen Problemen behaftet, da die Dokumente in unterschiedlichsten Formaten vorhanden sind, die Hard- und Software sich im ständigen Wandel befindet und die Entwicklungen in der Informationstechnologie immer rasanter fortschreiten. Die Standardisierung ist hier von großer Bedeutung. In dieser Diplomarbeit möchte ich zeigen, welche technischen und inhaltlich-organisatorischen Lösungsansätze bis jetzt im Bibliotheksbereich entwickelt wurden.

¹ vgl. Payer 2002a, S.2

2 Der Auftrag

Die UNESCO Charta zum Erhalt des digitalen Kulturerbes (UNESCO Charter on the Preservation of the Digital Heritage) wurde entwickelt, da weltweit immer mehr digitale Dokumente produziert werden, Dokumente, die ausschließlich digital hergestellt werden und zur Verfügung stehen und von Relevanz sind. Sie reichen von kulturellen, bildungserzieherischen, wissenschaftlichen, öffentlichen und administrativen Ressourcen bis hin zu technischen und medizinischen Informationen. Es ist wichtig, sobald wie möglich zu handeln um dieses digitale Kulturgut für jetzige und künftige Generationen zu erhalten.

Die Charta besteht aus 12 Artikeln, die Richtlinien und Maßnahmen für Bibliotheken enthalten, um die richtige Handhabung des digitalen Kulturerbes zu gewährleisten.

Hier nun ein Überblick über die wesentlichen Aussagen in diesen Artikeln:

Artikel 1 zeigt die wesentlichen Bereiche auf, in denen zunehmend digitale Ressourcen hergestellt werden (s. o.), diese digitalen Dokumente liegen in den unterschiedlichsten Formaten vor und in allen Teilen der Welt, deshalb ist hier ein internationales Management gefordert

Artikel 2 besagt, dass diese Dokumente für alle zugänglich gemacht werden sollen, deshalb ist es wünschenswert, dass die Mitgliedsstaaten mit relevanten Organisationen und Einrichtungen zusammenarbeiten, um den Zugang mit möglichst wenigen Beschränkungen zu gewährleisten und internationale Normen zu schaffen

Artikel 3 verdeutlicht den drohenden Informationsverlust, der durch mangelnde finanzielle Mittel, die rasante Entwicklung der Informationstechnologie und durch die fehlende gesetzliche Unterstützung eintreten wird

Artikel 4 ruft zur Notwendigkeit des Handelns auf und appelliert an das Bewusstsein aller, die Dringlichkeit des schnellen Handelns zu erkennen

Artikel 5 hebt hervor, wie wichtig es ist, ein System aufzubauen um die Langzeitverfügbarkeit und Authentizität der digitalen Dokumente zu gewährleisten

Artikel 6 fordert dazu auf gemeinsame Strategien und Verfahrensweisen zu entwickeln unter Einbeziehung der gegebenen Umstände, hierbei wird auch die Mitwirkung derer gefordert, die das Copyright in ihren Händen halten

Artikel 7 besagt, dass die Auswahlkriterien der Länder für das Sammelgut variieren können, jedoch auf einer definierten Verfahrensweise beruhen sollten

Artikel 8 empfiehlt die Entwicklung einer nationalen Politik um die Langzeitverfügbarkeit des digitalen kulturellen Erbes zu gewährleisten, so sollen gesetzliche und technische Rahmenbedingungen von der Erschließung bis hin zur Authentizität der Dokumente geschaffen werden

Artikel 9 verdeutlicht, dass das digitale Kulturerbe schon jetzt unermesslich ist, es ist zwar jeweils spezifisch für seine Kultur, jedoch wichtig um das Wissen über alle Kulturen zu erhalten

Artikel 10 betont, dass es möglich ist die Rollen und die Verantwortung bezüglich der Aufgaben zu verteilen; dieser Appell richtet sich an alle Beteiligten von Autoren, Softwareentwicklern bis hin zu den Verlagen, sie werden dazu aufgerufen mit den Nationalbibliotheken, Archiven und Museen zusammenzuarbeiten, private wie auch öffentliche Einrichtungen sollen sich mit einbringen.

Artikel 11 fordert das Eingehen von Partnerschaften aller Beteiligten, auch international, um die digitale Kluft zu überwinden und allen Ländern die gleiche Chance zu bieten ihr digitales Kulturerbe zu erhalten

Artikel 12 umschreibt die Rolle der UNESCO, diese fungiert auf nationaler und internationaler Ebene und sie wird alle Handlungen verfolgen und unterstützen um das Ziel dieser Charta zu erreichen, vorerst geht es um einen Zeitraum von 6 Jahren

(vgl. UNESCO 2003a).

Diese Charta ist der Auftrag zur Langzeitarchivierung von digitalen Dokumenten. Sie ist ein Aufruf für alle Beteiligten sich dieser Herausforderung zu stellen. Sie appelliert an das Bewusstsein aller sich der Aufgabe der Erhaltung des digitalen Kulturgutes anzunehmen, um einen Informationsverlust zu verhindern.

Insbesondere stellt diese Charta eine große Herausforderung für die National- und Regionalbibliotheken dar, da diese ihrem Sammelauftrag nachkommen müssen, aber genauso richtet sie sich auch an Autoren, Verlage und alle anderen, die im Prozess des digitalen Publizierens involviert sind. Die Bibliotheken und Archive sind sich des Problems schon länger bewusst und versuchen durch verschiedene Projekte Lösungen zu finden.

3 Digitale Dokumente

Dokumente aus dem WWW sind digitale Dokumente.

Was versteht man unter dem Begriff *Dokument* und was sind *digitale Dokumente*?

Der Terminus *Dokument* wird nach Kampffmeyer und Rogalla folgendermaßen definiert:

"Ein Dokument ist durch verschiedene Merkmale gekennzeichnet:

- physische Eigenschaften (Papier, Datei u.ä.)
- formale Eigenschaften (Aufbau, Gestaltung u.ä.)
- Ordnung (fachliche Zugehörigkeit, Reihenfolge, Version u.ä.)
- Inhalt (inhaltlicher Bezug u.ä.)
- Charakter (Archivierungswürdigkeit, Rechtscharakter, Bearbeitungsmöglichkeiten u.ä.)
- Zeit (Erzeugungsdatum, Verfallsdatum, letzte Benutzung, Aufbewahrungsfrist u.ä.)
- Erzeuger (Absender, Ersteller u.ä.)
- Nutzer (Empfänger, berechtigter Bearbeiter, Leser, letzter Bearbeiter

u.ä.)"

(Kampffmeyer und Rogalla 1997, S. 3).

Ein Dokument ist also als ein Objekt anzusehen, das an Hand seiner Merkmale zu erkennen ist. Die Merkmale hier unterliegen eher inhaltlich-formalen Kriterien. Das rührt daher, dass ein Dokument ursprünglich meist in Papierform vorlag. Ein Dokument war oftmals mit einer Urkunde gleich zusetzen.

Im Zeitalter der digitalen Medien liegen Dokumente nicht nur zunehmend in digitaler Form vor, sondern häufig auch ausschließlich in digitaler Form vor.

Die Definition für ein *digitales Dokument* lautet nach Endres und Fellner:

"Ein digitales Dokument ist eine in sich abgeschlossene Informationseinheit, deren Inhalt digital codiert und auf einem elektronischen Datenträger gespeichert ist, so daß er mittels eines Rechners genutzt werden kann."

(Endres und Fellner 2000, S. 15)

Diese Definition berücksichtigt ausschließlich die technische Seite, es werden keine inhaltlich-formalen Kriterien genannt. Weiterhin schreiben Endres und Fellner, dass der Begriff Dokument hier nicht als ein Dokument im eigentlichen Sinne anzusehen ist. Sie machen zwei Ausnahmen, einmal für *digitale Betriebsmittel*², d.h. wenn Betriebsmittel im Zusammenhang mit dem Internet inbegriffen sind und andererseits für *digitale Objekte*, hiermit sind Objekte gemeint, die selbständig keinen Wert haben, wie z. B. ein Inhaltsverzeichnis oder ein Abbildungsverzeichnis (vgl. Endres und Fellner 2000, S. 15).

Seit 2003 gibt es noch eine Erklärung, was unter *digitalen Dokumenten* zu verstehen ist. Sie ist von Borghoff, Rödiger, Scheffczyk und Schmitz, Informatiker der Universität der Bundeswehr München, die sich speziell mit der Problematik der Langzeitarchivierung befassen und auch das gleichnamige Buch dazu geschrieben haben, sie lautet:

"Ein *digitales Dokument* besteht aus dem digitalen Dokumenteninhalt und den zum Dokument gehörigen Metadaten. Der *digitale Dokumenteninhalt* ist als ein Strom (Folge) von Zeichen eines endlichen Zeichensatzes A codiert. Zu den Metadaten gehört die Angabe des zur Codierung des Dokumenteninhalts

² Dies leiten sie aus dem engl. ab von *digital resource*

verwendeten Zeichensatzes A sowie weitere Angaben (u. a. Autor, Titel). Die Metadaten sind als Folgen eines (Basis-)Zeichensatzes B codiert. Die Zeichensätze A und B können, müssen aber nicht übereinstimmen. Die Struktur der Metadaten und der zugrunde liegende Zeichensatz B sind archivabhängig festgelegt."

(Borghoff u. a. 2003, S.11)

Hier ist eindeutig zu erkennen, dass die Begriffsbestimmung für digitale Dokumente im Zusammenhang mit der Methodik der Langzeitarchivierung erfolgt, da hier gleich ein Bezug zu den Metadaten hergestellt wird, die in der Langzeitarchivierung eine wichtige Rolle spielen.

Es spielt also bei der Begriffsbestimmung eine große Rolle aus welchem Blickwinkel sie erfolgt, für Endres und Fellner war es wichtig ein digitales Dokument in Bezug auf eine digitale Bibliothek zu definieren, bei Borghoff und Kollegen wird ein digitales Dokument gleich in Zusammenhang mit der elektronischen Archivierung gebracht, bei der es relevant ist die Metadaten mit einzubringen.

Auch in der Literatur, in der englischen wie in der deutschen, werden viele verschiedene Begriffe verwendet, hinter denen sich die *digitalen Dokumente* verbergen, hier einige Beispiele:

digitale bzw. elektronische Ressourcen, digitale bzw. elektronische Publikationen, elektronische Dokumente, digitale bzw. elektronische Quellen oder auch *digitale Objekte*

In dieser Diplomarbeit geht es ausschließlich um digitale Dokumente aus dem WWW, d. h. Dokumente, die im Internet über das Hyper Text Transfer Protocol zur Verfügung stehen. Diese können in den unterschiedlichsten Dateiformaten vorliegen.

Diese digitalen Dokumente werden auch als Netzpublikationen bezeichnet, das können z. B. sein:

- elektronische Zeitschriften
- elektronische Hochschulschriften

- thematische Websites wie sie in der UNESCO Charta aufgeführt sind
- Online-Datenbanken
- Abbildungen, Ton, Video
- retrodigitalisierte Dokumente...

3.1 Gängige Dateiformate und ihre Besonderheiten

Digitale Dokumente liegen in den unterschiedlichsten Dateiformaten vor. Über das Datenformat lässt sich das Dateiformat definieren.

Hierzu eine Erklärung von Borghoff und Kollegen:

"Ein Datenformat heißt *binär*, wenn die zugehörigen Dokumenteninhalte nur Folgen von Nullen und Einsen sein können. Wenn der Dokumenteninhalt aus Zeichen eines anderen Zeichensatzes (z. B. dem des ASCII-Codes) besteht, dann spricht man von einem *textbasierten* Datenformat."

(vgl. Borghoff u. a. 2003, S. 12)

Aufgrund dieser unterschiedlichen Art der Kodierung lassen sich Dokumente in Text-Dokumente und Nicht-Text- Dokumente trennen (vgl. Zheng 2003, S. 2).

Ein Überblick über die gängigsten Dateiformate:

3.1.1 Text-Dokumente

3.1.1.1 PS (PostScript)

PS wurde von der Firma Adobe entwickelt, es ist eine Seitenbeschreibungssprache und legt somit das Format einer Seite fest, es besteht hier die Möglichkeit Kommentare einzufügen, die später nicht angezeigt werden. Um das PS-Format ausdrucken zu können benötigt man einen PostScript-fähigen Drucker (vgl. Winkler 2004, S. 639).

3.1.1.2 PDF (Portable Document Format)

Das PDF wurde auch von der Firma Adobe entwickelt und arbeitet ebenso mit

einer Seitenbeschreibungssprache. Innerhalb einer PDF-Datei sind alle Elemente komprimiert abgelegt, deshalb wird der Acrobat Reader o. ä. für die Darstellung benötigt. Dieser ist kostenlos zu erhalten. Zur Erstellung von PDF-Dateien wird ein spezielles Programm benötigt, dies ist kostenpflichtig und lizenziert. Eine Nachbearbeitung solcher Dateien ist nicht mehr möglich. PDF-Dateien sind geräteunabhängig (vgl. Ostermann 1999, S. 31f.).

3.1.1.3 DOC (Word-Dokument)

Word-Dokumente werden mit dem Textverarbeitungsprogramm WinWord von der Firma Microsoft erstellt. Für das Programm werden Lizenzen vergeben. Es bietet eine breite Palette an Funktionen an, mit deren Hilfe das Layout des Dokumentes bestimmt werden kann. Ein großer Nachteil von Word ist, dass bei der Anwendung unterschiedlicher Word-Versionen Kompatibilitätsprobleme auftreten können. Ein Vorteil ist, dass es inzwischen schon möglich ist den Text über Sprachsteuerung einzugeben (vgl. Winkler 2004, S. 810).

3.1.1.4 LaTeX

LaTeX ist ein Textsatzsystem. Es wird besonders im naturwissenschaftlichen Bereich genutzt, da es für den Umgang mit vielen mathematischen Formeln geeignet ist. LaTeX ist eine Weiterentwicklung von Tex, das von Donald Knuth entwickelt wurde. Mit Hilfe eines Textfiles wird das Layout eines Dokuments bestimmt (vgl. Winkler 2004, S. 459).

3.1.1.5 SGML (Standard Generalized Markup Language)

SGML ist standardisiert und gilt als "Urmutter" der Metasprachen (ISO 8879), sie legt die logische Struktur und den Inhalt der Dokumente fest. Sie wird für umfangreiche Dokumente verwendet, da es später noch möglich ist nur den Inhalt zu filtern und weiter zu verarbeiten. Deshalb eignet sich SGML sehr gut für den Datenaustausch. SGML ist jedoch eine recht komplexe Beschreibungssprache bzw. Metasprache und deshalb nicht sehr verbreitet. SGML ist plattformunabhängig. Derivate sind HTML und XML (vgl. Winkler 2004, S. 737).

3.1.1.6 XML (Extensible Markup Language)

XML ist ein Standard, der aus SGML entwickelt wurde. XML ist weniger komplex und deshalb als einfache Variante von SGML anzusehen. XML ist eine Metasprache, mit der sich Document Typ Definitions (DTDs) erstellen lassen, durch die DTD werden die Formatparameter festgelegt und somit ein bestimmter Dokumententyp definiert. Im Gegensatz zu HTML (Hyper Text Markup Language) ist XML erweiterbar. Da sich in DTDs, die mit XML erstellt wurden, Inhalt und Layout trennen lassen, ist XML für den Datenaustausch geeignet. Wie SGML ist XML nicht an bestimmte Software oder Hersteller gebunden (vgl. Extensible Markup Language 2004).

3.1.1.7 HTML (Hyper Text Markup Language)

HTML ist eine mit SGML erstellte DTD, in diesem Falle eine DTD für die Erstellung von Websites. HTML ist nicht erweiterbar, deshalb ist es auch nicht möglich in HTML eigene Befehle zu definieren. Befehle, die der Browser nicht kennt, zeigt er nicht an. Zudem variiert die Darstellung von HTML-Dokumenten von Browser zu Browser (vgl. Ostermann 1999, S. 33).

3.1.2 Nicht-Text-Dokumente

3.1.2.1 JPEG (Joint Photographic Experts Group)

Das JPEG-Format wurde von der Joint Photographic Experts Group entwickelt. Es ist zur Anzeige und Speicherung von Fotos- und Halbtonbildern geeignet. Pro Datei ist es möglich ein Bild einzubinden. Die Farbtiefe liegt bei 24 Bit. JPEG-Bilder können über einen WWW-Browser angezeigt werden. JPEG ist auch ein Standard für ein Kompressionsverfahren. Dieses Verfahren ist zwar verlustbehaftet, jedoch kann der Nutzer auf die Kompression selbst Einfluss nehmen (vgl. Winkler 2004).

3.1.2.2 GIF (Graphics Interchange Format)

Dieses Format wurde von der Firma Unisys entwickelt. Es ist fähig mehrere Bilder pro Datei zu speichern. Außerdem weist es multimediale Fähigkeiten auf, wie beispielsweise das Anzeigen von animierten Bildsequenzen. Das GIF-Format

reduziert die Darstellung auf 256 Farben. Die neuere Version des GIF-Formats GIF89a lässt Interlacing zu, d. h. die Bildzeilen werden nach und nach geladen, folglich wird das Bild schon während der Ladezeit z. T. angezeigt. Eine verlustfreie Datenkompression ist hier möglich. Aufgrund seiner reduzierten Farbdarstellung wird das GIF-Format eher für flächige Bilder und nicht für Fotos verwendet (vgl. Winkler 2004).

3.1.2.3 PNG (Portable Network Graphics)

Das PNG-Format war quasi die Antwort auf das GIF-Format als die Firma Unisys Lizenzanforderungen stellte. Entwickelt wurde es von der PNG Development Group des W3C (World Wide Web Consortium). Es bietet eine Farbtiefe von 24 Bit, das ist eine Verbesserung gegenüber dem GIF-Format. Eine verlustfreie Kompression ist möglich, was es wieder in Konkurrenz zum JPEG-Format stellt. PNG bietet gleichzeitig wie GIF Interlacing, jedoch wird das PNG-Format bis jetzt nur von neueren Browsern unterstützt (vgl. Winkler 2004).

3.1.2.4 MP3 (MPEG-1/2 Audio Layer-3)

MP3 wurde vom Fraunhofer Institut für integrierte Schaltungen in Erlangen in Kooperation mit AT&T Bell Labs und Thomson entwickelt. Es ist ein Verfahren zur Kompression von Audiodateien, bei welchem Töne, die für das menschliche Gehör nicht wahrnehmbar sind reduziert werden. MP3 ist ein verlustbehaftetes Verfahren, dies ist stark abhängig von der Übertragungsfrequenz. Um MP3-Dateien abspielen zu können wird ein spezieller MP3-Player benötigt (vgl. MP3 2004).

Anhand dieses Überblicks, und das sind längst nicht alle Dateiformate, die es gibt, kann man erkennen, dass die jeweiligen Dateiformate die unterschiedlichsten Besonderheiten aufweisen. Außerdem kommen meist mehrere Dateiformate zusammen, beispielsweise auf einer Website gibt es meist Bilder, einen Download-Bereich, dahinter verbergen sich dann oftmals PDF-Dateien, und noch viele weitere Elemente. Die Dokumente liegen vorwiegend in hybrider Form vor. Dies sind alles Aspekte, die die Eignung eines digitalen Dokumentes für die Langzeitarchivierung beeinflussen, da jede Datei einzeln archiviert werden muss. Hierbei sollte eine

Konvertierung der Dateien weitestgehend vermieden werden, da jede Konvertierung einen Datenverlust mit sich bringen kann und dadurch die Authentizität des Dokuments nicht mehr gewährleistet werden kann. Zusätzlich sind viele Formate in ihrer Entwicklung noch nicht abgeschlossen. Manche dieser Dateiformate können auch nur innerhalb proprietärer³ Systeme erstellt und dargestellt werden, die Technologie schreitet hier so schnell voran, dass die Software immer schneller veraltet und die Dokumente oftmals nicht mehr lesbar sind. Um all diese Probleme zu umgehen, ist es wichtig hier den Weg zu einer Standardisierung zu finden, dafür scheinen SGML und XML im Moment die beste Basis zu sein (vgl. Ostermann 1999, S.34; vgl. Dobratz u. a. 2001).

3.2 Dokumentenstandards

Um den Prozess der Langzeitarchivierung vereinfachen zu können ist es wichtig Standards zu schaffen. Diese Standards werden für technische als auch für organisatorische Aspekte der Langzeitarchivierung benötigt. Dies beginnt bei der Erstellung der Dokumente, es wäre wünschenswert, dass schon hierbei Standards eingehalten werden, einerseits um die inhaltliche Dokumentenstruktur zu bewahren, andererseits um Migrations- und Emulationskosten zu sparen (siehe Kapitel 6). Daher müssen für gängige Dokumentenformate DTDs geschaffen werden (Workshop 2002, S. 22). Dieses Ziel stellt hohe Anforderungen an die Autoren der Dokumente, daher ist es wichtig, diese in Zukunft durch Schulungen zu unterstützen. Auf diesem Wege könnte eine neue Kultur des elektronischen Publizierens geschaffen werden (Workshop 2002, S. 22f.). Noch sind wir weit von diesem Szenario entfernt, es gibt zwar Schulungen⁴ an diversen Hochschulen, jedoch immer noch für die gängigen Dateiformate wie Word und Latex, als Präsentationssoftware wird in der Regel PDF gewünscht. Hier besteht noch großer Handlungsbedarf.

Bei der Entwicklung von neuen Standards sollte man sich nicht auf Firmenstandards festlegen: Vielmehr scheint hier die Lösung bei der Nutzung der Markup-Sprachen zu liegen.

³ proprietär hier im Sinne von: an lizenzierte Software gebunden

⁴ Anmerkung, Bsp. <http://web.urz.uni-heidelberg.de/Dokumentation/Kurse.html>

Ein Datenformat, das als Standardformat anzusehen ist, sollte nach Borghoff u. a. 2003, S. 42 folgenden Anforderungen Stand halten:

- das Datenformat sollte offen gelegt sein
- das Datenformat sollte von einer internationalen Standardisierungsinstitution als Standard ausgezeichnet sein, z. B. ISO, W3C,... oder wie z. B. das PDF-Format als De-facto-Standard, d. h. von der Öffentlichkeit, anerkannt sein
- das Datenformat sollte ein häufig verwendetes sein um den aktuellen Hard- und Softwareanforderungen zu entsprechen
- das Datenformat sollte nicht lizenziert und frei von Patentansprüchen sein

Diese Kriterien erfüllen bis jetzt nur SGML und XML, die zum Erstellen von Text-Dokumenten verwendet werden, somit besteht noch großer Handlungsbedarf bei der Schaffung von Standards für Nicht-Text-Dokumente (vgl. Dobratz u. a. 2001). Bei letzteren kommt als weitere Problematik noch der Qualitätsverlust durch die Konversion hinzu.

Auf der Basis von SGML und XML werden durch das W3C weitere Standards entwickelt wie z. B. XSL (Extensible Style Language) und XLink (Extensible Linking Language). Mit Hilfe dieser Metasprachen ist es möglich DTDs zu entwickeln wie z. B. xDML für *Dissertationen Online*⁵, MathML für mathematische Formeln, CML für chemische Formeln oder auch der Open e-Book Standard. Diese Standards werden vorab in Pilotprojekten getestet und gegebenenfalls weiterentwickelt werden (vgl. Dobratz und Tappenbeck 2002).

3.3 Vor- und Nachteile von digitalen Dokumenten

Die stetig steigende Anzahl von Websites zeigt wie beliebt digitale Dokumente sind. Im September 2002 waren über die Suchmaschine *google*⁶ 2.468.940.685 Websites (Engster 2002, S. 10) recherchierbar, jetzt im Jahre 2004 ist schon eine "Suche auf 4.285.199.774 Web-Seiten" möglich, wie auf der Startseite von *google* angezeigt wird⁷.

⁵ <http://www.dissonline.de>

⁶ <http://www.google.de>

⁷ Zugriff am: 2004-07-31

OCLC (Online Computer Library Center) brachte bis einschließlich 2002 innerhalb des Projekts *Web Characterization Project*⁸ eine jährliche Statistik über den Zuwachs von Websites heraus, leider gibt es für 2003 keine neue Statistik. Die Übersicht über die Jahre 1998 bis 2002 zeigt uns den rasanten Anstieg der bis dahin stattgefunden hat:

Zeitraum	Anzahl	Wachstum
1998-1999	4.882.000	71%
1999-2000	7.399.000	52%
2000-2001	8.745.000	18%
2001-2002	9.040.000	3%

Tabelle 1: Jährliches Wachstum von Websites (OCLC 2002)

Wie aus der Tabelle zu ersehen ist, lässt der "Websit-Boom" jetzt wieder nach, wahrscheinlich weil die meisten Firmen, Institutionen, Privatpersonen usw. ihren Bedarf gedeckt haben.

Deshalb ein Überblick über die Vor- und Nachteile digitaler Dokumente:

Vorteile:

Speicherplatz digitale Dokumente brauchen im Gegensatz zu Printpublikationen wesentlich weniger Speicherplatz

Schnelligkeit wenn man ein digitales Dokument lokalisiert hat, ist es sehr schnell verfügbar, es lässt sich sofort übertragen, man muss nicht erst den Weg zur Bibliothek auf sich nehmen

⁸ <http://www.wcp.oclc.org>

Gleichzeitige Nutzung ein digitales Dokument steht mehreren Personen gleichzeitig zur Verfügung, vorausgesetzt es ist copyright-frei, in einer Bibliothek kann es passieren, dass das Medium verliehen ist

Selektive Informationsvermittlung Printpublikationen stehen oft nur in Form einer größeren Einheit zur Verfügung, z.B. eine Zeitschrift erscheint periodisch und hat ein bestimmtes Volumen, wenn man nun nur einen bestimmten Artikel braucht, muss man trotzdem auf das ganze Heft zurückgreifen. Bei einer Zeitschrift, die online zur Verfügung steht, ist es möglich nur diesen Artikel anzuschauen

Weltweite Verfügbarkeit bei digitalen Dokumenten ist es egal, an welchem Ort sie zur Verfügung stehen, sie sind jederzeit von überall her abrufbar

Weiterverarbeitbarkeit digitale Dokumente sind meist erweiterbar, d.h. sie können weiterbearbeitet und genutzt werden

Erschließbarkeit digitale Dokumente bieten unterschiedliche Möglichkeiten der Erschließung, z.B. über Metadaten oder über Freitextsuche. Auch dem Nutzer stehen diverse Möglichkeiten zur Verfügung wie die akustische Wiedergabe oder auch die Darstellung in Blindenschrift

Einbindung verschiedener Medien in digitale Dokumente können nicht nur Text, sondern auch Ton, Bild, Video eingebunden werden

Nachteile:

Abhängigkeit von technischen Hilfsmitteln um digitale Dokumente erstellen und nutzen zu können werden technische Hilfsmittel benötigt, dies erfordert einen finanziellen Hintergrund und technisches Know-how

Leichte Veränderbarkeit digitale Dokumente können verändert werden, ohne dass dies erkennbar ist, deshalb müssen bestimmte Sicherheitsvorkehrungen getroffen werden

Datenverlust digitale Dokumente sind auf einem Trägermedium gespeichert. Wird dieses Trägermedium beschädigt, kann es zu einem Datenverlust kommen, dadurch kann das Dokument beschädigt werden oder auch gar nicht mehr zur

Verfügung stehen. Datenverlust kann auch durch Datenkonversion entstehen (vgl. Kapitel 6)

Übertragung und Sicherheit die Benutzung des WWW ist immer mit Gefahren verbunden, deshalb ist es wichtig auf die Sicherheit zu achten

Aufwand für Langzeitarchivierung die Langzeitarchivierung von digitalen Dokumenten erfordert eine Anpassung an die Fortschritte der Informationstechnologie, sonst kann es passieren, dass die Dokumente nicht mehr lesbar sind. Deshalb ist der Prozess der Langzeitarchivierung auch sehr kostenaufwendig

Hier muss man hinzufügen, dass die Tendenz zu digitalen Dokumenten abhängig ist von den Fachgebieten, sie werden z. B. gerne für die Naturwissenschaften und die Medizin genutzt, während in den Geisteswissenschaften immer noch Papier bevorzugt wird (vgl. Endres und Fellner 2000, S. 15ff.).

Oft wird behauptet, dass durch die Nutzung digitaler Dokumente Papier eingespart wird, aber aus der allgemeinen Erfahrung heraus, kann man sagen, dass die Dokumente, wenn länger mit ihnen gearbeitet wird, trotzdem ausgedruckt werden, da es so einfacher ist sie zu lesen, von daher lässt sich diese These nicht belegen.

4 Organisatorische Konzepte

Die Langzeitarchivierung von digitalen Dokumenten stellt hohe Anforderungen an ein Archivsystem. Das Ziel ist nicht nur die Dokumente langfristig zu archivieren, sondern sie auch langfristig nutzbar zu machen. Deshalb sollte ein digitales Archiv folgende Kriterien erfüllen:

Langfristige Benutzbarkeit

Benutzbarkeit es sollte eine kontinuierliche Benutzbarkeit gewährleistet sein, die Daten sollten jederzeit interpretierbar und lesbar sein

Authentizität es muss garantiert werden können, dass das Dokument hier in seiner Originalfassung vorliegt und dass der Verfasser der echte Verfasser ist

Integrität und Funktionalität hiermit ist die Information gemeint, die besagt wie die Dokumente miteinander verknüpft sind und welche Funktionalitäten eingebettet sind

(vgl. ÖNB 2003).

Um diesen Anforderungen gerecht zu werden hat die RLG/OCLC Working Group einen Leitfaden für *Verlässliche Digitale Archive – Trusted Digital Repositories* erarbeitet, in dem die wesentlichen Eigenschaften für ein digitales Archiv genannt werden. Weitere wichtige Eigenschaften, zu den oben schon genannten, sind (RLG/OCLC 2002):

- das OAIS-RM (Open Archival Information System Referenz Model) soll Grundlage für das Archivsystem sein
- es sollen organisatorische und administrative Strukturen vorhanden sein
- es soll in einem angemessenen finanziellen Rahmen umgesetzt werden
- in dem System sollen hardware- und softwaretechnische Anforderungen angepasst werden
- die Sicherheit des Systems soll gewährleistet sein
- alle Unternehmungen und Verfahren sollen transparent sein

Einen Teil dieser Kriterien finden wir in der UNESCO-Charta. Hinzu kommt, dass das OAIS-RM zum Einsatz kommen soll. Dieser Entschluss wurde innerhalb des Projekts NEDLIB⁹ (European Networked Deposit Library) gefasst. Teilnehmer dieses Projekts waren acht europäische Nationalbibliotheken, u. a. Die Deutsche Bibliothek, ein Nationalarchiv, zwei Softwarefirmen und drei wissenschaftliche Verlage. Durchgeführt wurde es von Januar 1998 bis Dezember 2000, Koordinierungsstelle war die Niederländische Nationalbibliothek, die Koninklijke Bibliotheek. Ziel des Projekts war es, Lösungen für die Langzeitarchivierung auf europäischer Ebene zu finden, auf der Basis eines DSEPs (Deposit System for Electronic Publications). Anfangs hatte man versucht ein eigenes Modell für das DSEP zu entwickeln, bis das OAIS-RM vorgeschlagen wurde. Dieses bot fast alle Funktionalitäten, die für die Langzeitarchivierung digitaler Dokumente benötigt

⁹ <http://www.kb.nl/coop/nedlib>

werden. Ein Prozess wurde von NEDLIB hinzugefügt, der Prozess *Preservation Planning*. Dies ist ein spezieller Prozess, der für die Langzeitarchivierung benötigt wird. In ihm sind alle Aufgaben definiert, um den Zugriff auf die Dokumente auch in Zukunft gewährleisten zu können, es geht hier z. B. um die Überwachung des Hardware- und Softwaremarkts oder um Migration und Emulation (siehe Kapitel 6).

Jene Funktionen, die nicht vom OAIS-RM abgedeckt werden und auch nicht vom traditionellen Bibliotheksmanagementsystem ausgeführt werden, wurden über das DSEP ergänzt (vgl. Liegmann 2001; vgl. Borghoff u. a. 2003, S. 35).

4.1 Das OAIS-Referenzmodell (OAIS-RM)

Das OAIS-RM ist ein logisch konzeptionelles Modell, das auf herkömmliche Archive und für die Archivierung digitaler Dokumente angewendet werden kann. Es dient zur Umsetzung von organisatorischen Aspekten und zum Teil auch von technischen Aspekten. Da es modular aufgebaut ist eignet es sich um die Archivierung in Kooperation planen zu können. Entwickelt wurde das OAIS-Referenzmodell ursprünglich vom *Consultative Committee for Space Data Systems (CCSDS)* im Auftrag der *International Organization for Standardization (ISO)*. Das CCSDS untersteht der NASA und wurde gegründet um den Datenaustausch in der Weltraumforschung zu verbessern. Im Februar 2003 ist das OAIS-RM zum Standard erklärt worden (ISO 14721:2003). Mit Hilfe des OAIS-Referenzmodells lassen sich die Entitäten und Abläufe bestimmen. Es gibt drei grundlegende Modelle.

4.1.1 Das Umgebungsmodell

In diesem Modell wird die Umgebung dargestellt, in der sich das OAIS-Prozessmodell (hier als OAIS archive bezeichnet) befindet, welches die eigentliche Aufgabe der Archivierung unter sich hat. Im Umgebungsmodell gibt es drei Schnittstellen.

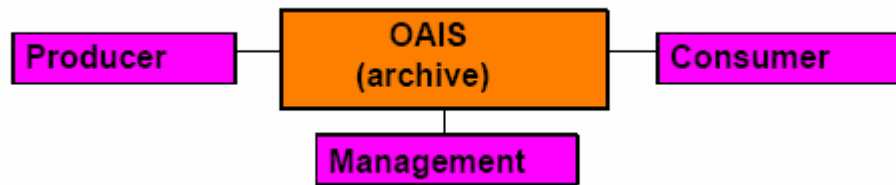


Abbildung 1: OAIS-Umgebungsmodell (CCSDS 2002b, S. 2)

zum Producer er ist der Erzeuger eines Dokuments, das können Autoren, Institutionen o. a. sein, sie sind diejenigen, die die Informationen zur Archivierung bereitstellen

zum Management zum Management zählt z. B. welche Informationen gesammelt werden sollen, die eigentliche Administration des Archivs zählt hier nicht dazu, sie ist in das OAIS-Prozessmodell integriert

Consumer der Consumer ist der Kunde, der die Dienste des OAIS letztendlich in Anspruch nimmt, weil er das Objekt betrachten möchte

(vgl. CCSDS 2002b, S. 2).

4.1.2 Das Informationsmodell

Vorab ist noch anzumerken, dass im OAIS-RM den Begriffen *Daten* (engl. Data Object) und *Informationen* (engl. Information Object) eine besondere Bedeutung zugeordnet wird. Daten, also in Form von Bits, sind noch nicht als Informationen anzusehen. Vielmehr müssen die Daten erst interpretiert werden, dies funktioniert nur, wenn die nötige *Knowledge Base* vorhanden ist. Unter Knowledge Base wird das nötige Wissen verstanden, um die Daten interpretieren zu können. Dieses Zusatzwissen wird in OAIS-RM als *Representation Information* bezeichnet (vgl. Borghoff u. a. 2003, S. 26f.).

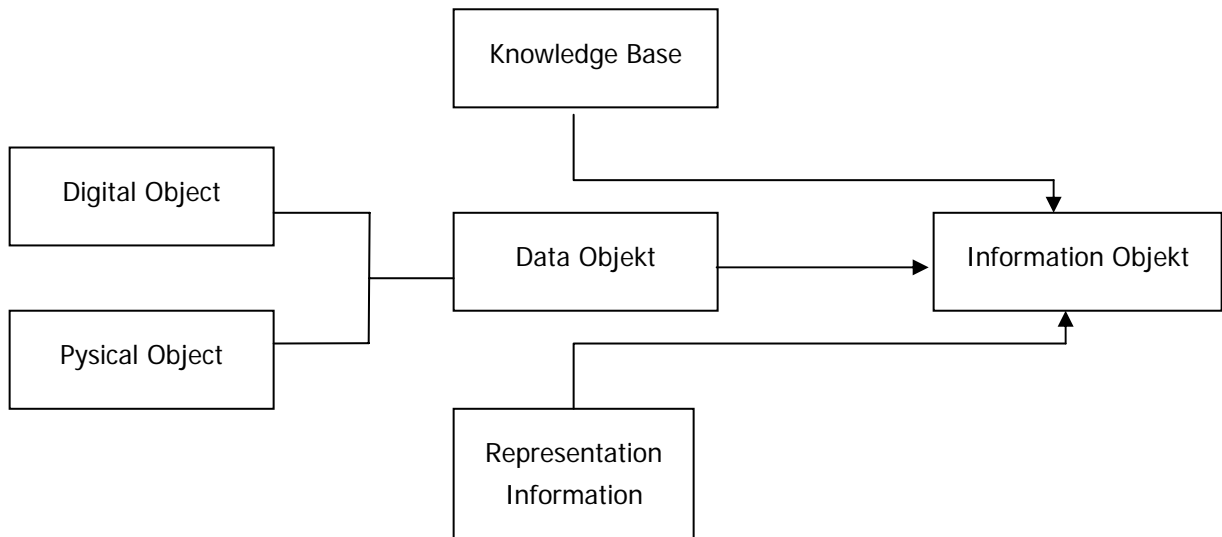


Abbildung 2: OAIS-Informationsmodell (Borghoff u. a. 2003, S. 27)

Diese Trennung des Verständnisses von Informationen und Daten ist wichtig für die Erläuterung des nächsten Begriffs, der im OAIS-RM eine zentrale Rolle spielt: das *Information Package*. Das Information Package setzt sich zusammen aus: *Content Information* und *Preservation Description Information (PDI)*.

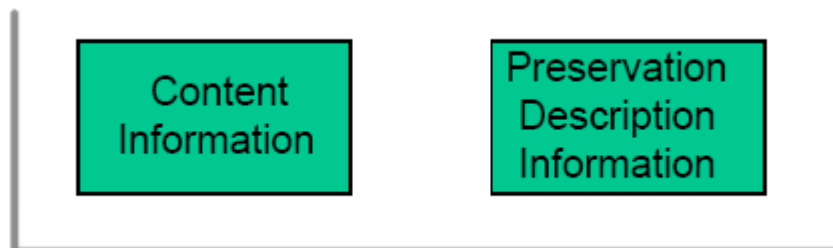


Abbildung 3: Information Package (CCSDS 2002b, S. 3)

Hinter der Content Information verbirgt sich das eigentliche Objekt, das es zu archivieren gilt und die Informationen, die für die Darstellung benötigt werden, die Representation Information, im OAIS-Informationsmodell mit Information Object benannt (vgl. OCLC/RLG 2001, S.13).

Beispiel:

Es soll eine Website archiviert werden, in dieser Website sind PDF-Files integriert, die es nun auch zu archivieren gilt. Dann ist dieses PDF-File als ein Data Object anzusehen, mit der zugehörigen Representation Information und der Knowledge Base, also zusätzliche Informationen, z. B. über die Adobe-Software mit der das PDF-File letztendlich angezeigt wird, ergibt sich hieraus das Information Object bzw. die Content Information (vgl. CCSDS 2002b, S. 3).

Nun ist schon ersichtlich, dass diese zusätzlichen Informationen sich nur auf das PDF-File beziehen, es fehlen noch weitere Informationen, die benötigt werden um das PDF-File in Zusammenhang mit der Website zu bringen, die entsprechenden Metadaten finden wir in der Preservation Description Information. Deshalb finden wir auch in der Erklärung von Borghoff und Kollegen in Kapitel 3 die Aussage, dass zu einem digitalen Dokument entsprechende Metadaten dazu gehören. Die PDI lässt sich noch in vier "Container" unterteilen:

Reference Information enthält Informationen über die Content Information bezüglich des Retrievals, z. B. die Suche mit Hilfe eines DOIs (Digital Object Identifier)

Context Information enthält Informationen über die Beziehungen zu anderen Objekten

Provenance Information enthält die jeweilige Änderungshistorie der Content Information

Fixity Information dient der Authentizitätssicherung z. B. durch die Vergabe von digitalen Signaturen

(OCLC/RLG 2001, S. 14).

Diese vier "Container" sind im Archivsystem separat verwaltet.

Jedes Information Package besitzt noch eine *Packaging Information*, diese stellt die Verbindung zum physischen Speichermedium her und jedes Information Package bekommt noch eigene Metadaten dazu, die im OAIS-RM als *Descriptive Information* bezeichnet werden. In den Metadaten werden Informationen

festgehalten, die das Wiederauffinden im Archiv ermöglichen (vgl. Borghoff u. a. 2003, S. 28).

Abhängig davon, wo sich nun ein Information Package im OAIS-Prozessmodell befindet, werden drei Arten von Information Packages definiert:

SIP (Submission Information Package) dieses wird vom Producer an das Archiv gesendet

AIP (Archival Information Package) ist das Information Package, welches im Archiv gespeichert wird

DIP (Dissemination Information Package) ist das Information Package, das der Consumer aus dem Archiv zur Betrachtung erhält

Dies ergibt folgendes Gesamtbild für ein AIP:

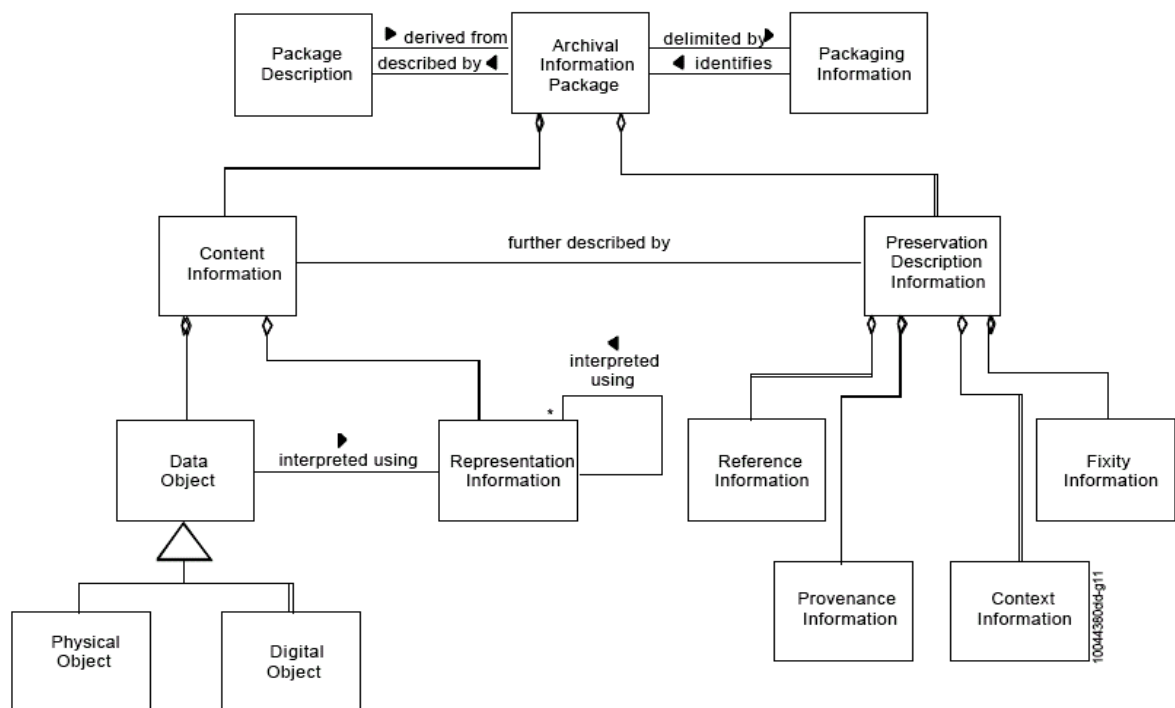


Abbildung 4: OAIS-Informationsmodell (OCLC/RLG 2001, S. 11)

Diese Information Packages sind als Eingabedaten für die Prozesse anzusehen, die im eigentlichen Archiv ablaufen. Dieses Archivsystem wird als OAIS-Prozessmodell dargestellt (vgl. Borghoff u. a. 2003, S. 29).

4.1.3 Das Prozessmodell

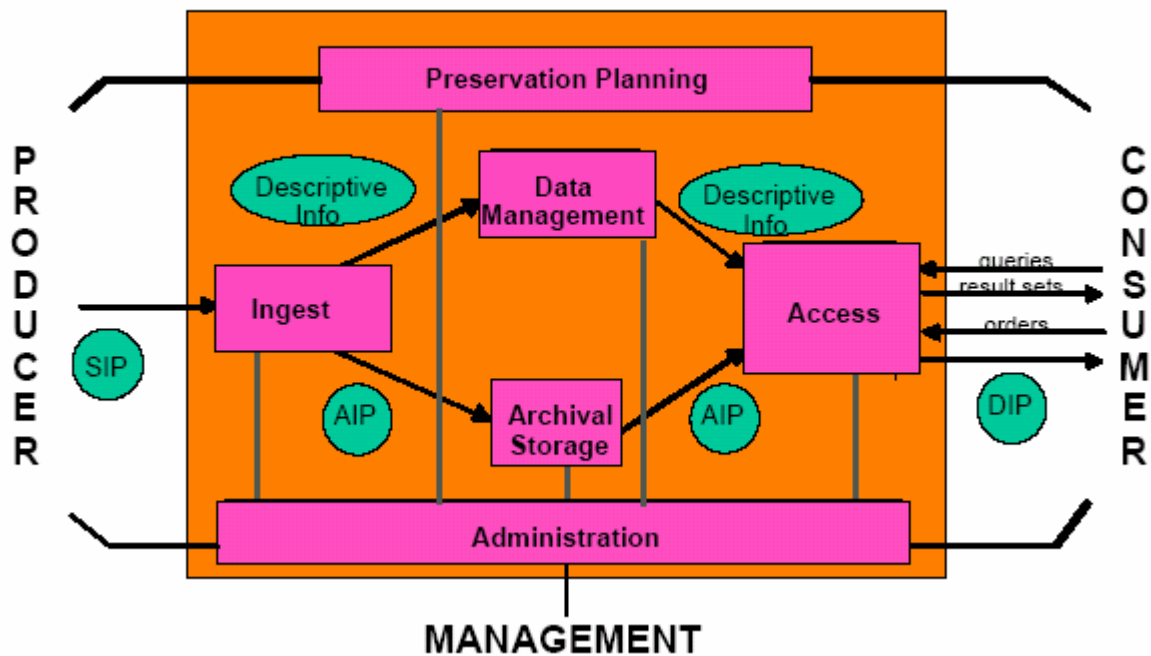



Abbildung 5: OAIS-Prozessmodell (CCSDS 2002b, S. 6)

Das Prozessmodell ist in sechs Prozesse unterteilt, die hier als  dargestellt sind, ihnen sind folgende Aufgaben zugewiesen:

Ingest (Eingangsbearbeitung) ist der erste Prozess, er ist die Schnittstelle zwischen dem Producer und dem OAIS-Prozessmodell, hier wird das eingehende SIP, das von einem Producer gesendet wird, übernommen und für die Speicherung und Verwaltung im Archiv vorbereitet. Das SIP wird durch den Ingest-Prozess kontrolliert, ob es alle Kriterien erfüllt wie z. B. Vertragsbedingungen,

Gewährleistung von Vollständigkeit und Authentizität, usw.

Wie dies genau ablaufen soll, lässt das OAIS-RM offen, deshalb ist es als ein logisch konzeptionelles Modell anzusehen.

Wird das SIP akzeptiert, kann es in ein AIP umgewandelt werden und wird durch die Descriptive Information ergänzt. Das AIP wird an den Prozess Archival Storage weitergeleitet und die Descriptive Information an den Prozess Data Management.

An diesem Punkt gelangen wir wieder zu dem Wunsch nach Dokumentenstandards. Optimal wäre es, wenn das Data Object, also das digitale Dokument, als ein XML-Schemata vorliegt, somit wäre es ein sogenanntes "archivierungsfreundliches" Format, da sich Inhalt und Layout gut trennen lassen ohne, dass die Representation Information auf eine bestimmte Hard- und Software angewiesen ist. Nur so ist es bis jetzt möglich Migrations- und Emulationskosten zu umgehen.

Archival Storage (Archivierung) ist der zweite Prozess, er nimmt die AIPs an und ist für ihre Speicherung zuständig. Ebenso gibt dieser Prozess die AIPs auf Anfrage weiter zu Access. Weitere Aufgaben von ihm sind die Unversehrtheit der AIPs zu garantieren und Backups durchzuführen.

Data Management (Datenmanagement) ist der dritte Prozess, er verwaltet die Descriptive Information der AIPs. Diese Daten werden in einer Datenbank gesammelt, das Data Management gibt sie auf Anfrage wieder heraus und bereitet sie auf.

Administration (Verwaltung) ist der vierte Prozess, er ist zuständig für die routinemäßige Verwaltung des Archivs. Die wichtigste Aufgabe des Administrations-Prozesses ist die Überwachung der Einhaltung von Standards, aber auch die Optimierung von Hardware- und Softwarekomponenten fällt in seinen Verantwortungsbereich.

Preservation Planning (Planung und Steuerung) ist der fünfte Prozess, er ist ein Prozess mit besonderer Bedeutung für die Langzeitarchivierung. Durch ihn soll der IT-Markt im Auge behalten werden um den rasanten Entwicklungen von Hard- und Software folgen zu können. Er soll auch entscheiden, welche Strategien

für die Langzeitarchivierung angewendet werden sollen.

Access (Objektbereitstellung) Access ist der sechste und letzte Prozess, er ist die Schnittstelle zwischen Consumer und dem OAIS-Prozessmodell. Über ihn kann der Consumer nach Dokumenten suchen z. B. mittels eines Online-Kataloges, wenn er fündig geworden ist, schickt er seine Suchanfrage ab und der Access-Prozess liefert das gewünschte DIP aus.

(vgl. Borghoff u. a. 2003, S. 29ff.; vgl. CCSDS 2002b; S. 2ff.)

Durch die Festlegung dieser Schlüsselprozesse und deren Schnittstellen in Form des OAIS-Referenzmodells ist die Infrastruktur des Archivsystems gegeben. Dies ist aber für die Langzeitarchivierung nicht ausreichend. Im Glossar des Workshops 2002, S.44 finden wir zum OAIS-Referenzmodell folgende Aussage:

"Das als ISO 14721 verabschiedete Referenzmodell beschreibt ein Archiv als Organisation, in dem Menschen und Systeme mit der Aufgabenstellung zusammenwirken, Informationen zu erhalten und einer definierten Nutzerschaft verfügbar zu machen. OAIS-Konformität beinhaltet die Übernahme der Verantwortung für die Einhaltung der im Referenzmodell vorgegebenen Rahmenbedingungen. Das Referenzmodell spezifiziert weder das Design noch die Implementierung eines OAIS-konformen Archivs.

[<http://www.ccsds.org/documents/pdf/CCSDS-650.0-B-1.pdf>]"

(Workshop 2002, S. 44)

Das OAIS-Referenzmodell ist als ein Leitfaden für die Organisation eines Archivs zur Langzeitarchivierung anzusehen, dessen Rahmenbedingungen eingehalten werden sollen, wie die Umsetzung erfolgt wird jedoch offen gelassen. Es wird natürlich versucht soviel wie möglich mit Hilfe von Computersystemen zu verwirklichen um eine gewisse Automatisierung der Prozesse zu erreichen.

In der weiteren Vorgehensweise werden deshalb die Schritte definiert, die nicht vom OAIS-Prozessmodell und vom Bibliotheksmanagementsystem abgedeckt werden.

4.2 Das DSEP-Prozessmodell

Die Begriffe OAIS-RM und DSEP-Prozessmodell beginnen nun zu verschmelzen, deshalb eine weitere Erläuterung aus dem Glossar des Workshops 2002, S.43 für den Begriff *Depotsystem*:

"...Im engeren Sinne die technische Implementierung eines OAIS-konformen Archivs. Im erweiterten Sinne gehören dazu auch Systeme außerhalb des OAIS-konformen Archivs:

- die vor der Eingangsschnittstelle (OAIS: ingest) liegende Infrastruktur zum Transfer der digitalen Objekte von den Produzenten in das Archiv und
- die hinter der Auslieferungsschnittstelle (OAIS: access) liegende Infrastruktur der Endnutzer-Umgebungen, in denen die digitalen Objekte benutzt werden sollen."

Diese zwei neuen Prozesse sind als Delivery & Capture und Packaging & Delivery benannt worden. Hier eine Darstellung der Prozesse eines DSEPs:

The top-level processes

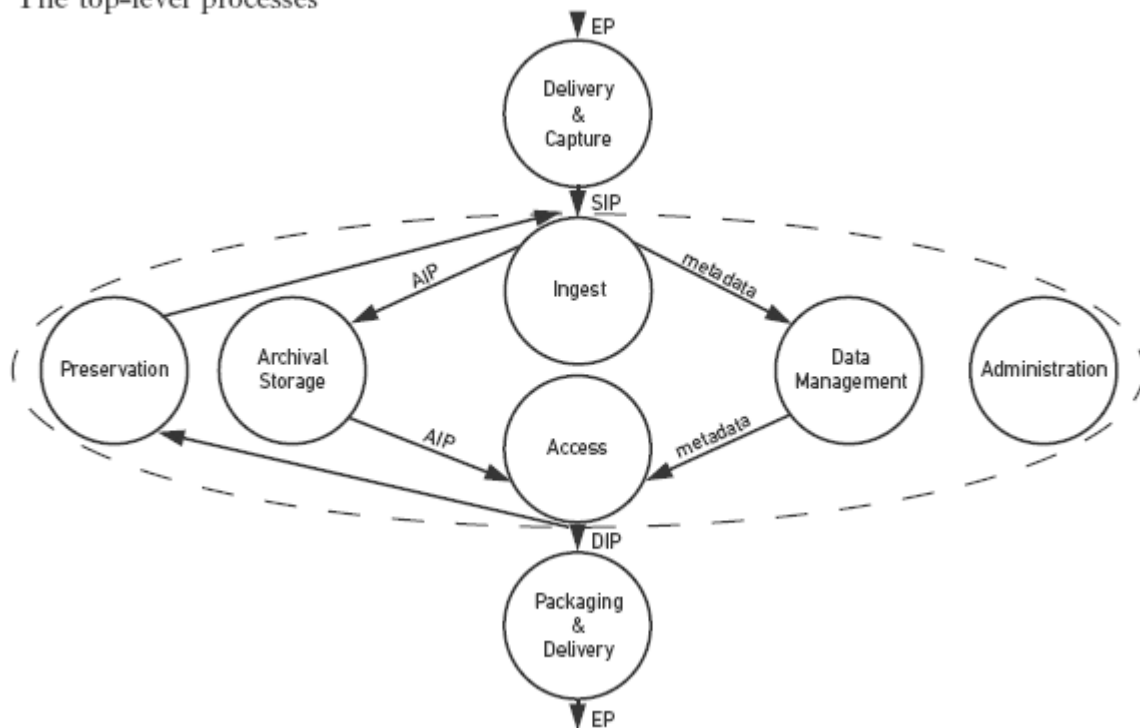


Abbildung 6: DSEP-Prozessmodell (van der Werf 2000, S. 21)

EP steht hier für "elektronische Publikationen".

Delivery & Capture hier gehen alle elektronischen Publikationen ein, die über das DSEP archiviert werden sollen. Über diesen Prozess werden sie in SIPs umgewandelt, die an Ingest weitergegeben werden, vorab werden sie überprüft, ob sie den Standards für ein DSEP entsprechen (vgl. Borghoff u. a. 2003, S. 34). NEDLIB hat hierzu ein Papier erarbeitet (NEDLIB Consortium 2000) in dem die entsprechenden Standards dargelegt sind.

Packaging & Delivery dieser Prozess liefert die gewünschten elektronischen Publikationen, die er über Access anfordert, an den Endnutzer (im OAIS-Umgebungsmodell als Consumer bezeichnet) aus. Access liefert das DIP, Packaging & Delivery entpackt das DIP und stellt es in entsprechender Form dar, so dass es vom Endnutzer betrachtet werden kann. In dem DIP können zusätzlich

Metadaten enthalten sein und Software, die zur Präsentation benötigt werden (vgl. Borghoff u. a. 2003, S. 34).

Zusätzlich wurde der Prozess Preservation Planning, der im OAIS-RM durch NEDLIB ergänzt wurde, über das DSEP noch weiter ausgebaut.

Er wird im DSEP *Preservation* genannt und in zwei Subprozesse unterteilt (vgl. Borghoff u. a. 2003, S. 35f.):

Preservation Planning gliedert sich in:

- *Develop Preservation Standards & Strategies* dieser Prozess ist zuständig für die Entwicklung und Konzeption von Standards
- *Develop Packaging Designs* soll über das Design der Information Packages entscheiden und deren Weiterentwicklung
- *Define Reference Platforms* macht Angaben zu Hardware- und Softwareanforderungen um eine Referenz-Plattform aufbauen zu können und das entsprechende Dokument betrachten zu können
- *Monitor Technologie* wird der Prozess genannt, der dazu dient, den IT-Markt zu überwachen

Preservation Activities ist unterteilt in:

- *Create New Reference Platforms* ist dazu da um die neu definierten Referenz-Plattformen nachzubilden und sie in Form von SIPs im Archiv zu speichern
- *Archival Information Update* dieser Prozess soll die Aktualisierung der AIPs gewährleisten

Da jetzt alle notwendigen Prozesse für die Langzeitarchivierung digitaler Dokumente durch das DSEP definiert sind und der Aufbau des DSEPs modular gestaltet ist, kann jetzt entschieden werden, welche Vorgänge vom Bibliotheksmanagementsystem ausgeführt werden können und für welche Vorgänge noch Lösungen zu finden sind. Der Geschäftsgang einer Bibliothek besteht aus drei Schritten:

- Erwerbung
- Katalogisierung
- Vermittlung/Benutzung

Diese werden heute über das Bibliotheksmanagementsystem abgewickelt. Für elektronische Publikationen sind jedoch zusätzliche Vorgänge einzuplanen. Im NEDLIB-Bericht Nr. 6 von van der Werf 2000 auf S. 7 sind die Anforderungen an eine Archivbibliothek im Umgang mit elektronischen Publikationen in 11 Schritten beschrieben:



Abbildung 7: Workflow des DSEP (Liegmann 2001)

(1) Auswahl

Für Printpublikationen gibt es bestimmte Auswahlkriterien, diese wird es ebenso für EPs geben. Das *Kompetenznetzwerk Langzeitarchivierung*¹⁰ (siehe Kapitel 8) arbeitet hierzu noch *Empfehlungen für Sammelrichtlinien und Auswahlkriterien für die Archivierung digitaler Quellen* aus (vgl. OBwald 2004). Die Auswahl erfolgt über die Erwerbsabteilung und wird unter Einsatz des Bibliotheksmanagementsystems abgewickelt.

¹⁰ <http://www.langzeitarchivierung.de>

(2) Erwerbung

Die EPs werden entweder über den herkömmlichen Weg bestellt und gekauft z. B. elektronische Zeitschriften oder der Autor gibt sie von sich aus an die Bibliothek z. B. einen wissenschaftlichen Bericht oder es können Netzpublikationen sein, die die Bibliothek durch ihren Sammelauftrag erhält. Hinzu kommen die Metadaten, außerdem noch die vertraglichen Bedingungen, falls welche vorhanden sind. Der letzte Punkt wird im OAIS-RM von Ingest ausgeführt, liegt aber praktischerweise in der Hand der Erwerbsabteilung. Dies wird ebenfalls mit dem Bibliotheksmanagementsystem ausgeführt und verwaltet (vgl. Borghoff u. a. S. 34f.).

(3) Transfer

Der Prozess Transfer entspricht dem Prozess Delivery & Capture, also ein Prozess der eindeutig dem DSEP zugeschrieben wird.

Eine Kopie der EP wird an das Archiv übermittelt. Hierzu gibt es noch keine einheitlichen Lösungen, vielmehr werden diese jetzt erst mit Hilfe von Projekten ermittelt. Dissertationen Online verwendet OAI-Schnittstelle und OAI-Protokolle¹¹ (vgl. Steinke 2004).

(4) Registrierung

Printpublikationen erhalten einen Besitzvermerk. Ein vergleichbares Verfahren ist für EPs erforderlich (vgl. Liegmann 2001).

(5) Verifikation

Bei der Verifikation geht es darum die Authentizität und Integrität der EPs nachzuweisen. Dies ist eine Aufgabe des DSEP, die sehr schwierig zu lösen ist, da die digitalen Dokumente zunehmend komplexer gestaltet sind, es wird hier versucht mit automatischen Verfahren zu arbeiten (vgl. Liegmann 2001).

¹¹ Diese wurden von der Open Archives Initiative (OAI) entwickelt, diese Initiative sucht nach Lösungen um freien Zugriff auf verteilte Archive, unter Einbeziehung von Metadaten, zu verbessern

(6) Katalogisierung

Die Katalogisierung ist ein Vorgang der auch für EPs über das Bibliotheksmanagementsystem abgewickelt wird. Hierzu können möglicherweise auch vorhandene deskriptive Metadaten genutzt werden.

(7) Speicherung

Die EP wird in das Speichersystem des DSEP aufgenommen. Hier wird sichergestellt, dass die EP lesbar bleibt, z. B. über Migration oder Emulation und die Backups werden hier durchgeführt. Dies entspricht dem Prozess Archival Storage im OAIS-RM. Wie die Umsetzung dieses Prozesses genau erfolgen soll, scheint noch offen zu sein.

(8) Langzeiterhaltung

Hinter der Langzeiterhaltung verbergen sich alle Maßnahmen, die für die Langzeitarchivierung digitaler Dokumente von Relevanz sind. Dies entspricht den Aufgaben, die im Prozess Preservation Planning im OAIS-RM definiert sind, im DSEP wurden sie noch feiner untergliedert. Auch dieser Prozess ist so komplex, dass es noch seine Zeit braucht um entsprechende Lösungen zu finden.

(9) Reaktivierung

Die Reaktivierung entspricht dem DSEP-Prozess Packaging & Delivery. Die Kopie des EP wird bereitgestellt. Für die Auslieferung wird nach einer möglichst standardisierten Form gesucht um die Anforderungen, die an die Umgebung gestellt werden, möglichst gering zu halten.

(10) Benutzung

Unter Benutzung wird die Umgebung (Bibliothek, Ausstattung und Service) verstanden, in welcher der Endnutzer die EP betrachtet und alle Optionen, die er benötigt um Zugriff auf die EP zu erhalten, wie z.B. User-Identifikation, Verwaltung der Zugriffsrechte, User-Profile usw.

In der neueren Literatur bei Borghoff und Kollegen (vgl. Borghoff u. a., S. 33) ist unter (9) die *Bestandsvermittlung*, unter (10) die *Auslieferung der EP* und unter Schritt (11) die *Zugangsvermittlung* verzeichnet, diese sind hier noch unter

Reaktivierung und Benutzung zusammengefasst.

(11) Systemüberwachung

Der Workflow des DSEP soll automatisch überwacht werden um die Qualität der EPs zu sichern und einen Informationsverlust zu verhindern.

Dieser Schritt ist bei Borghoff und Kollegen unter Punkt (12) zu finden.

5 Weitere organisatorische Aspekte

In Kapitel 4 wurden die Abläufe und Entitäten dargelegt, die zur Organisation eines Archivs benötigt werden. In diesem Kapitel werden die weiteren Punkte beschrieben, die bei der Langzeitarchivierung von Netzpublikationen eine Rolle spielen.

5.1 Erschließung von Netzpublikationen

Netzpublikationen müssen wie Printpublikationen erschlossen werden um sie dem Nutzer zugänglich zu machen. Deshalb ist es notwendig eine formale wie auch eine inhaltliche Erschließung durchzuführen.

Die Formalerschließung für Print-Medien in wissenschaftlichen Bibliotheken erfolgt über RAK-WB (Regelwerk für die alphabetische Katalogisierung für wissenschaftliche Bibliotheken), für die Formalerschließung von Netzpublikationen ist RAK-NBM (Regelwerk für die alphabetische Katalogisierung für Nicht-Buch-Materialien) vorgesehen. Diese Regelwerke sind nationale Regelwerke, die auf Basis der ISBD (International Standard Bibliographic Description) ausgearbeitet wurden. Die inhaltliche Erschließung erfolgt über die RSWK (Regeln für den Schlagwortkatalog), hier wird mit kontrollierten Termini gearbeitet und unter Anwendung einer Klassifikation.

Die Deutsche Bibliothek (DDB) hat im Rahmen ihres Projekts Dissertationen Online mit RAK-NBM und mit Metadaten, in Form von Dublin Core (DC), gearbeitet. Die Aufnahme von Titeldaten mit RAK-NBM hat sich jedoch als unzureichend, bezüglich

Indexierung und Retrieval, und als sehr aufwendig erwiesen (vgl. BSB 2000, S. 4). Hier scheint die Lösung beim Einsatz von Dublin Core zu liegen, DC deckt die formale und die inhaltliche Erschließung ab und bietet weitere Elemente, die für die Erschließung von Netzpublikationen relevant sind. Außerdem hat die Anwendung von DC folgende Vorteile:

- die Metadaten können vom Autor selbst über ein Formular bearbeitet werden, hierzu ist kein bibliothekarisches Fachwissen wie bei RAK gefordert (vgl. BSB 2000 S. 3)
- die Metadaten können direkt in das Dokument eingebunden werden, dies ist bei einem HTML-Dokument unter dem Befehl <META NAME...> möglich (vgl. BSB 2000 S. 8)
- die Metadaten müssen nicht in das Dokument eingebettet sein, sie können auch verlinkt werden und somit erst im nachhinein erstellt werden (vgl. BSB 2000 S. 6)

Dublin Core wurde von der *Dublin Core Metadata Initiative (DCMI)*¹² entwickelt. Das *Dublin Core Metadata Element Set (DCMES)* besteht aus 15 Elementen (ausführlich in Kapitel 5.2), sie dienen dazu, das digitale Dokument einheitlich zu beschreiben. Die DCMI will damit folgende Ziele erreichen:

- Einfache Erstellung und Wartung
- eine allgemein verständliche Bedeutung, so dass jeder damit umgehen kann
- internationale Gültigkeit
- Erweiterbarkeit

(WU 2002a).

Um Dublin Core weiter zu spezifizieren wurden *Qualifier* entwickelt. Sie werden angewendet um die Bedeutung der Elemente verfeinern zu können (WU 2002b) und entsprechen somit eher den Anforderungen, die Bibliothekare an die Erschließung stellen.

DC ist kein Regelwerk, d. h. es ist nicht festgelegt in welcher Form die Daten einzugeben sind, sie sind nicht normgerecht (BSB 2000), somit liegt die

¹² <http://www.dublincore.org>

intellektuelle Nachbearbeitung der Daten bei der Katalogisierungsabteilung der Bibliothek, daher soll hier möglichst schnell agiert werden, die DDB schreibt hierzu (DDB 2004c):

"Für die Erschließung elektronischer Ressourcen sollen Regeln entwickelt und als Standard eingeführt werden, die die Nutzung von Metadaten einbeziehen, international abgestimmt sind und in die Erschließung der Print-Ressourcen integriert werden können."

Über das Projekt META-LIB, das von 1998-2002 andauerte, hatte man schon versucht sich entsprechenden Lösungen anzunähern. Die Arbeitsstelle für Standardisierung der DDB arbeitet jetzt an der Metadaten-Standardisierung weiter.

5.2 Metadaten

Metadaten spielen bei der Langzeitarchivierung eine große Rolle, sie haben die Aufgabe den kompletten "Lebenszyklus" eines digitalen Dokuments zu beschreiben. Hierzu sind verschiedene Typen von Metadaten erforderlich:

Content Metadata sind bibliographische Metadaten, sie werden auch deskriptive Metadaten genannt

Structural Metadata sind für die Navigation im Archivsystem erforderlich

Administrative Metadata sind Metadaten, die für die verteilte Archivierung und die Pflege der Archivsystems eine Rolle spielen

Technical Metadata enthalten technische Daten, sie machen z. B. Angaben über die Dateigröße oder die technischen Daten, die durch eine Digitalisierung herbeigeführt werden

Preservation Metadata enthalten spezielle Daten, die für die Langzeitarchivierung benötigt werden, wie das Format des archivierten Dokumentes und die Software-Informationen, die zur Darstellung benötigt werden

Metadata about Terms and Conditions hier sind alle Angaben zum Copyright und zum Urheberrecht enthalten

...

(vgl. Neuroth 2002, S. 3f.)

Das Projekt Carmen (CARMEN steht für Content Analysis, Retrieval, Metadata: Effective Networking)¹³, behandelte im AP (Arbeitspaket) 2/5¹⁴ das Thema *Metadaten zu Terms und Conditions und zur Archivierung*. AP 2/5 lief von 1.10.99 bis zum 28.02.2002, während dieser Zeit wurden Prototypen für Metadateninformationen entwickelt. Die Konzeption der Prototypen basierte weitgehend auf Dublin Core (vgl. Carmen 2000). Im OAIS-RM finden wir die Metadaten unter der Preservation Description Information, dort sind sie untergliedert in vier "Container" (siehe Kapitel 4.1.2). Carmen AP 2/5 hat versucht die Metadatentypen auf die Definitionen dieser vier "Container", soweit als möglich, umzulegen:

Reference Information ist gleichzusetzen mit Content Metadata (deskriptive Metadaten) und entspricht den Elementen von Dublin Core (Version 1.1), bis auf Source (entfällt, da hier von einer reinen Netzpublikation ausgegangen wird) und Relation (siehe Context Information).

Hieraus ergeben sich folgende Angaben:

- title (enthält den Titel des Dokuments)
- creator (zeigt den Urheber des Dokuments an)
- subject (beinhaltet das Thema des Dokumentes, in Form von Stichwörtern oder Schlagwörtern)
- description (entspricht einem Abstract, aber auch als Table of Contents (ToC) möglich)
- publisher (zeigt den Verleger oder den Anbieter des Dokumentes an)
- contributor (hier werden weitere Urheber verzeichnet oder Personen, die zu diesem Dokument etwas beigetragen haben)
- date (Datum)
- type (gibt die Art des Dokuments an, im Sinne von Vortrag, Report,...)
- format (macht Angaben über das Dateiformat, wie z. B. HTML, PDF,...)

¹³ <http://www.mathematik.uni-osnabrueck.de/projects/carmen/>

¹⁴ <http://www.sub.uni-goettingen.de/carmen/>

- identifier (enthält die eindeutige Identifikation des Dokuments z. B. URL, URN, DOI)
- language (Sprache)
- coverage (beinhaltet Angaben über zeitliche oder geographische Gegebenheiten)
- rights (enthält Aussagen zum Urheberrecht)

(vgl. DCMI 2003)

- *reference metadata identifier*

hinzu kommt der jeweilige Identifier für jeden Container der PDI.

Context Information hier sind die Felder verzeichnet, die ursprünglich unter dem Element Dublin Core *Relation* als Qualifier aufgelistet sind, dies entspricht folgenden Angaben:

- Is Version Of/Has Version
- Is Replaced By/Replaces
- Is Referenced By/References
- Is Required By/Requires
- Is Format Of/Has Format
- Is Part Of /Has Part
- *context metadata identifier*

Provenance Information zeigt die Änderungshistorie der Content Information aus dem OAIS-RM an, hierunter sind folgende Angaben zu finden:

- Original Media Type
- Original Technical Environment
- Ingest Process History
- Preservation History
- Rights Management

- *provenance metadata identifier*

Fixity Information enthält Angaben zur Authentizitätssicherung des Dokuments, eine Dokumentation dieses Vorgangs, sowie die digitale Signatur, Prüfsummen, Zeitstempel usw. und einen *fixity metadata identifier*

Wie zu sehen ist konnte teilweise Dublin Core integriert werden (in Reference Information und Context Information), teilweise sind die Lösungen noch offen und es stellt sich die Frage, ob es möglich ist, weitere Core Sets zu entwickeln, zusätzlich müssen die "Container" noch genauer gegeneinander abgegrenzt werden und es ist ratsam Richtlinien zu erarbeiten. Ungewiss ist noch, ob die Möglichkeit besteht, Metadaten automatisch zu generieren. Bei der Schaffung von Standards für die Interoperabilität gibt es weitere offene Punkte, wie z. B. welches Encoding Schema (XML, RDF¹⁵,...) angewendet werden soll oder wie sich die Architektur gestalten soll (dezentral, zentral, hybrid) u. v. a., bei dieser Thematik werden zu vielen Fragen erst noch Lösungen gesucht (vgl. Neuroth 2002, S. 5ff.)

Auf internationaler Ebene sind viele Arbeitsgruppen aktiv, die sich mit der Untersuchung der *Preservation Metadata* befassen (hierunter ist nicht der Typus Preservation Metadata zu verstehen, sondern alle Metadaten, die für die Langzeitarchivierung benötigt werden, dies können auch deskriptive oder administrative Metadaten sein), da dies ein komplexes noch unerforschtes Gebiet ist. Beispiele hierfür sind die OCLC/Working Group on Preservation Metadata, die National Library of New Zealand (NLNZ), die National Library of Australia (NLA) und die Harvard University. Auf europäischer Ebene befassten sich die Projekte NEDLIB und CEDARS (Curl exemplars in digital archives)¹⁶, letzteres lief von 1998 bis 2002, mit diesem Thema. Über CEDARS und NEDLIB, wie auch von der NLA und der Harvard University wurden Metadatenelementsets ausgearbeitet. OCLC/RLG hat diese in einem White Paper zusammengestellt. Es hat den Anschein, dass zur Zeit noch diskutiert wird welches Set sich am besten für die Implementierung eignet (vgl. OCLC/RLG 2001, S. 37).

¹⁵ RDF steht für Resource Description Framework und ist ein vom W3C empfohlenes Metadatenformat <http://www.w3.org/RDF/>

¹⁶ <http://www.leeds.ac.uk/cedars/>

5.3 Authentizität und Sicherheit

Um die Authentizität und Integrität von digitalen Dokumenten gewährleisten zu können sind Sicherheitsmaßnahmen erforderlich. Es muss belegt werden, dass das Dokument im Original vom benannten Autor vorliegt und nicht im Nachhinein manipuliert wurde. Dies ist eine Anforderung, der ein digitales Archiv nachkommen muss, um als vertrauenswürdiges Archiv gelten zu können (vgl. Kapitel 4, S. 22/23).

Der Einsatz von Sicherheitsmaßnahmen beginnt bei der Hardware. Hier ist eine Trennung von Dokumenten- und Archivserver notwendig. Der Dokumentenserver sollte zertifiziert sein um seine Identität sicherzustellen und für den Nutzer überprüfbar zu sein. Der Übertragungsweg sollte per SSL (Secure Sockets Protokoll) gesichert sein (vgl. Scholze 2002, S. 24).

Zusätzlich müssen die Dokumente geschützt werden, entweder durch die Bildung von Prüfsummen oder durch die Vergabe von Signaturen.

Die Ermittlung der Prüfsumme wird auch als *Hashing* bezeichnet. Mittels Hashing wird eine Zahl berechnet, die ein Dokument eindeutig identifiziert, die Prüfsumme oder der Hashcode. Ändert sich etwas im Dokument wird auch der Hashcode ein anderer. Der Hashcode muss beim Übertragen eines Dokumentes geschützt sein, deshalb sollte er verschlüsselt sein. In den Hashcode wird zusätzlich ein Schlüssel miteingebracht, dieser ist nur dem Empfänger und dem Sender bekannt (vgl. Hashing 2002), dies erfolgt meist mittels MD-5 (Message Digest Algorithm, Version 5), dieser Algorithmus ist inzwischen Standard für die Prüfsummenbildung, RFC 1321 (vgl. Engster 2003, S. 18).

Eine andere Lösung um die Authentizität und Integrität digitaler Dokumente sicherzustellen liegt in der digitalen Signatur. Die digitale Signatur ist inzwischen einer handschriftlichen Unterschrift rechtlich gleichgestellt. Ein Vorteil der digitalen Signatur ist, dass der Sender identifizierbar ist. Angewendet wird das Private-/Public-Key-Verfahren, das mittels bestimmter Programme umgesetzt wird, wie z. B. GPG (GNU Privacy Guard) oder PGP (Pretty Good Privacy) (vgl. Engster 2002, S.

18). Die digitale Signatur wird mittels asymmetrischer Kryptographie hergestellt. Asymmetrisch deshalb, weil zum Verschlüsseln (Sender) ein nichtgeheimer Schlüssel, der Public Key, und zum Entschlüsseln (Empfänger) des Dokuments ein geheimer Schlüssel, der Private Key, benötigt wird. Zusätzlich kann der Sender mit seinem Private Key das Dokument signieren, so kann der Empfänger den Sender anhand "seiner Unterschrift" eindeutig zuordnen, dazu ist es erforderlich, dass der Schlüssel vorher von einer Zertifizierungsstelle mit einem Zertifikat ausgestattet wurde. Zusätzlich arbeitet das Verfahren noch mit der Vergabe von Prüfsummen (vgl. Payer 2003, 6B.8.2.2. PGP).

Problematisch hierbei ist, dass digitale Signaturen ein Ablaufdatum haben, da Zertifikate nur für einen bestimmten Zeitraum ausgestellt werden. Deshalb ist es erforderlich die Dokumente mit einem Zeitstempel zu versehen, dieser wird über einen Zeitstempeldienst vergeben und gewährt, dass das Dokument nach diesem Zeitpunkt nicht mehr verändert wurde, so lässt sich eine Verifikation der Signatur vornehmen. Allerdings sind auch die Zeitstempel befristet und sollten regelmäßig geprüft und gegebenenfalls erneuert werden.

Um dieses Problem zu lösen wurde *XML Advanced Electronic Signatures (XadES)* entwickelt. Dies ist ein Standard der European Telecommunications Standards Institute (ETSI), in diesem Standard ist ein Datenformat definiert, das als digitale Langzeitsignatur anzusehen ist, durch ihn kann die Archivierbarkeit der Signaturen voraussichtlich sichergestellt werden. Allerdings ist dieser Standard in der Praxis noch nicht erprobt, dies wird wohl frühestens Ende des Jahres stattfinden (vgl. Egger 2003, S. 2ff.).

5.4 Persistent Identifiers – URN

Um eine Netzpublikation dauerhaft archivieren zu können muss sie eindeutig identifizierbar sein und eine dauerhafte Adressierung vorweisen. URLs (Uniform Resource Locator) sind für dieses Unterfangen zu instabil. Es braucht sich beim Verfasser der Website z. B. nur ergeben, dass er den Provider über den er die Internetseite zur Verfügung stellt, wechselt oder, dass die Institution, bei der er arbeitet, den Namen ändert und sich somit auch die Internetadresse ändert, ...

hierfür kann es sehr viele Gründe geben (vgl. Payer 2002a). Für den Nutzer ist die Website letztendlich nicht mehr aufrufbar und der Browser zeigt eine Fehlermeldung an. Ebenso geschieht dies bei Netzwerkfehlern, dadurch sind URLs temporär nicht ansprechbar.

Aufgrund dieser Problematiken sieht man die Lösung in der Anwendung eines Persistent Identifiers (PI).

Dies hat den Vorteil, dass eine strikte Trennung von Identifikation (weltweit eindeutig, persistent) und Standortreferenz gegeben ist (vgl. Schröder 2004a).

Hierzu eine Definition des Kompetenznetzwerkes Langzeitarchivierung zu PIs:

"Standortunabhängige, eindeutige Bezeichner für digitale Objekte; analog den ISBN-Nummern für Printpublikationen. Mit PIs wird ein stabiler Zugriff auf das digitale Objekt gewährleistet, wodurch dieses dauerhaft zitiert werden kann." (vgl. Kompetenznetzwerk Langzeitarchivierung 2004a)

OCLC hat einst versucht die Problematik der URLs (häufige Änderungen) durch den Einsatz von PURLs (Persistent Uniform Resource Locator) zu lösen. Ein PURL wird an einen Resolutionsdienst weitergeleitet, der über den PURL-Server die korrekte URL vermittelt., somit codiert ein RURL den Ort eines dazwischen liegenden Resolutionsdienstes. OCLC sah damals wohl keine Konsensfähigkeit und hat die Vergabe von PURLs als Zwischenlösung angesehen bis ein Standardverfahren gefunden wird (vgl. Borghoff u. a. 2003, S. 104).

Im Verlagswesen wird mit DOIs und Handle-Systemen gearbeitet. Über den DOI kann das Dokument eindeutig identifiziert werden, das Handle-System wird für die Namensresolution eingesetzt. DOIs wurden entwickelt um den Rechtsschutz in offenen verteilten Netzen zu gewährleisten (vgl. Borghoff u. a. 2003, S. 107).

Die DDB versucht jetzt die URL-Problematik durch die Vergabe von URNs zu lösen, hierzu wurde das Projekt EPICUR (Enhancement of Persistent Identifier Services – Comprehensive Method for unequivocal Resource Identification)¹⁷ ins Leben gerufen. Der Vorgänger von EPICUR war Carmen AP 4, in AP 4 wurde das Thema *Persistent Identifiers and Metadata Management in Science* bearbeitet, es wurde ein PI-Prototyp entwickelt und es wurde ein kooperatives URN-Management für

¹⁷ <http://www.persistent-identifier.de>

Online-Hochschulschriften eingeführt (vgl. Schröder 2004a), das in der DDB seit September 2001 umgesetzt wird (vgl. Diekmann u. a. 2001).

EPICUR übernahm die Migration des Prototypen und ist für die URN-Strategie Der Deutschen Bibliothek zuständig. Ziel des Projekts EPICUR ist es den PI-Dienst weiterzuentwickeln um die Anwendung verschiedener PI-Systeme, z. B. DOI und URN, gleichzeitig zu zulassen. Weiterhin soll eine Infrastruktur innerhalb Deutschlands für PIs aufgebaut werden und international wird die Zusammenarbeit mit der Nationalbibliothek Finnlands angestrebt um ein URN Information Centre einzurichten. Die Laufzeit ist von 2002 bis 2004 angesetzt (vgl. DDB 2004b, S. 3).

Die URN-Strategie der Deutschen Bibliothek beinhaltet folgende Aspekte:

- (1) Vergabe von URN-Unternehmensräumen
- (2) URN-Vergabe
- (3) URN-Registrierung
- (4) URL-Pflege
- (5) URN-Resolving
- (6) technische Dienstleistung der Deutschen Bibliothek

(vgl. Schröder 2004a).

Der Aufbau einer URN gliedert sich folgendermaßen:

URN:NID (Namespace):SNID (Subnamespace):NISS (Namespace Specific String)

Um internationale Zusammenarbeit gewährleisten zu können haben sich europäische und internationale Nationalbibliotheken mittels einer Initiative für den NID *NBN* entschieden. *NBN* steht für *National Bibliography Number* (vgl. Diekmann u. a. 2001).

Hieraus ergibt sich folgendes Bild URN:NBN:de für die DDB.

zu (1) urn:nbn:de [Bibliothekverband]:[Bibliothekssigel]-[eindeutige Produktionsnummer][Prüfziffer] für Institutionen, die an den Bibliotheksverbänden teilnehmen.

Bsp.: urn:nbn:de:gbv:089-3321752945 ist die URN einer Hochschulschrift an der

TIB Hannover, die Prüfziffer ist an dieser Stelle nicht ersichtlich

Für Institutionen, die nicht an den Verbünden teilnehmen: urn:nbn:de [vierstellige Ziffer]-[eindeutige Produktionsnummer][Prüfziffer]

zu (2) erfolgt für alle Netzpublikationen, die an der DDB archiviert werden und kann für Netzpublikationen, die auf zertifizierten Servern liegen, durchgeführt werden

zu (3) bei der Abgabe einer Netzpublikation ist der erste Schritt die Anmeldung dieser unter: http://www.ddb.de/cgi-bin/epmeld_form.pl. Im zweiten Schritt erfolgt die URN-Erstmeldung über <http://nbn-resolving.de/login.php>. Bei beiden Schritten muss die entsprechende URL, unter der das Dokument zu finden ist, jeweils mit angegeben werden.

zu (4) die DDB führt einen regelmäßigen Linkcheck der URLs auf ihre Verfügbarkeit durch, dieser Vorgang erfolgt automatisiert (vgl. Diekmann u. a. 2001)

zu (5) die DDB stellt einen zentralen Resolving-Dienst zur Verfügung, in ihm ist ein Verwaltungssystem integriert, das die URN mit zugehöriger URL, sowie alle institutions- und personengebundenen administrativen Daten speichert. Aufgabe des Resolving-Dienstes ist es, den Nutzer der ein Dokument mit der Angabe einer URN sucht automatisch an die gültige URL und an das entsprechende Dokument weiterzuleiten.

zu (6) die DDB unterstützt dies mit allen erforderlichen technischen Mitteln und versucht die Vorgänge zu optimieren

(vgl. Schröder 2004a).

Die DDB hat 2001 mit der Vergabe von URNs begonnen, seitdem steigt die Zahl stetig:

Jahr	Anzahl URNs
2001	158
2002	2880
2003	7827
2004-08-06	10427

Tabelle 2: Jährliches Wachstum der URNs (EPICUR 2004b)

Um einem Ausfall durch Netzwerkprobleme entgegen zu wirken wird im Bibliotheksservice-Zentrum (BSZ) in Konstanz ein Mirror-Server aufgebaut (vgl. EPICUR 2004a).

6 Strategien der Langzeitarchivierung

Um Netzpublikationen dauerhaft verfügbar zu halten bedarf es bestimmter Maßnahmen, da sich Hard- und Software im ständigen Wandel befinden. Bei der Archivierung von Netzpublikationen spielt das Trägermedium eine untergeordnete Rolle, da dies in der Regel eine Festplatte ist, hier spielt eher das Format eine Rolle, in dem das Dokument vorliegt, da zur Ansicht der Dokumente Programme benötigt werden, die das Format interpretieren können. Handelt es sich bei dem Format um eines, das über proprietäre Systeme erstellt wurde, kann die Lebensdauer unter Umständen nur wenige Jahre betragen, aber auch bei Standardformaten können sich im Laufe der Zeit Änderungen ergeben. Wird hier nicht rechtzeitig agiert und das Dokument nicht in ein neues Format übertragen, kann es passieren, dass es nicht mehr lesbar ist.

Um dieser Problematik entgegen zu treten gibt es mehrere Möglichkeiten (vgl. Borghoff u. a. 2003, S. 14ff.):

6.1 Das "Computer Museum"

Bei dieser Variante werden nicht nur die digitalen Dokumente archiviert, sondern auch die jeweilige Hard- und Software. Diese muss in Stand gehalten werden und es werden Fachleute benötigt, die damit umgehen können (vgl. Payer 2002a). Da die Hardware aber auch nur eine begrenzte Lebensdauer hat und die Software nach einer gewissen Zeit vom Markt verschwindet und nicht mehr erhältlich ist, ist dies eine Variante, die für die Langzeitarchivierung nicht von Bedeutung ist. Zudem kann der Nutzer die Dokumente auch nur vor Ort anschauen (vgl. ÖNB 2003, S. 3). Einziger Vorteil ist, dass die Authentizität des Dokuments gewährleistet wäre.

6.2 Migration

Bei der Migration oder Datenkonversion werden die digitalen Dokumente von Zeit zu Zeit in ein neueres Format übertragen. Die Migration ist die Archivierungsstrategie, die in der Praxis zum Teil schon angewendet wird. Hierzu gab es eine Befragung, die von Margaret Hedstrom und Sheon Montgomery im Auftrag der Research Libraries Group durchgeführt wurde. Es wurden 54 Mitgliedsinstitutionen befragt, von diesen gaben nur 18 an eine Migrationsstrategie zu verfolgen und nur 7 Institutionen hatten dies auch in schriftlicher Form festgehalten. Dies zeigt, dass der Ansatz zum Teil schon da ist, es in der Praxis aber noch viel zu wenig umgesetzt wird (vgl. Bischoff 2003).

Ein großer Nachteil der Migration ist, dass es durch die Konversion zu einem Datenverlust kommen kann. So kann es passieren, dass die Darstellung des Dokuments nicht mehr identisch ist mit der Darstellung des Originaldokuments oder noch schlimmer, dass es zu einer Änderung des Inhalts kommt, das kann z. B. bei einem Gedicht passieren, welches in Versform geschrieben ist (vgl. Neumann 2003, S. 55). Ein einfaches Beispiel ist die Konversion eines Dokuments, das mit Word 97 geschrieben wurde, nach Word XP. Hier treten speziell bei der Formatierung große Probleme auf. Wie sich die Veränderungen auf das zukünftige Dokument auswirken ist meist vorher nicht abschätzbar.

Um eine zuverlässige Migrationsstrategie umsetzen zu können ist ein sogenannter *Technology Watch* erforderlich. Dies bedeutet, die archivierende Institution muss die Entwicklungen auf dem IT-Markt ständig im Auge behalten (im OAIS-Modell wird diese Aufgabe dem Prozess *Preservation Planning* zugeordnet). Falls sich entsprechende Änderungen auf dem Hard- oder Softwaremarkt ergeben kann so rechtzeitig agiert werden (vgl. Neumann 2003, S. 54). Bei der rasanten Entwicklung, die auf dem IT-Markt stattfindet, bedeutet dies auch eine steigende Anzahl von Migrationsvorgängen, was wiederum die Kosten der Institution ansteigen lässt, welche einem sinkenden Etat gegenüber stehen (vgl. Neumann 2003, S. 56). Hinzu kommt die steigende Zahl der Datenbestände (vgl. Bischoff 2003).

Besonders problematisch ist die Anwendung der Migration bei Netzpublikationen, die meist als HTML-Dokumente vorliegen, in welche die unterschiedlichsten Dateiformate (MP3, JPEG, PDF,...) eingebunden sind, hinzu kommen noch die Verlinkungen. Hier steigt der Aufwand einer Migration immens und die Gefahr des Qualitätsverlusts bei der Darstellung steigt beträchtlich (vgl. Bischoff 2003).

Eine Möglichkeit der Migration soweit als möglich auszuweichen ist der Einsatz von Standardformaten, dadurch lassen sich die Migrationsintervalle verlängern, da standardisierte Formate im Gegensatz zu anderen eine längere Lebensdauer aufweisen (vgl. Borghoff u. a. 2003, S. 41f.).

Die Migration bildet gegenüber der Emulation folgende Vorteile:

- das Dokument liegt in einem aktuellen Dateiformat vor
- dementsprechend gibt es Werkzeuge mit dem das Dokument weiter verarbeitet werden kann
- ein aktuelles Dateiformat bietet in der Regel Mehrwert gegenüber einem älteren Format

(vgl. Borghoff u. a. 2003, S. 55f.)

6.3 Emulation

Unter Emulation, im Sinne der Langzeitarchivierung, versteht man die Nachahmung alter Systeme auf neuen Systemen. Um dies umsetzen zu können wird ein Emulator benötigt, dies ist ein Programm, mittels dessen die Hard- und Softwareumgebung des alten Systems auf dem neuen Rechner simuliert wird (vgl. Bischoff 2003). Es ist also notwendig für jede Quellplattform einen Emulator zu entwickeln. Dies ist ein großer Aufwand, aber im Gegensatz zu dem Aufwand, der bei der Anwendung von Migration betrieben wird, ein Schritt, der nur einmalig ausgeführt werden muss. Hinzu kommt, dass die Dokumente weiter in ihrer Originalversion betrachtet werden könnten und die Authentizität somit sichergestellt wäre. Dem gegenüber stehen der Aufwand und die Kosten, die durch die Erwerbung von Lizenzen für die entsprechende Software, entstehen (vgl. Borghoff u. a. 2003, S. 19).

Die Emulation ist ein Verfahren, das bislang noch nicht in Archiven eingesetzt wird, sondern für die Entwicklung neuer Hardware angewendet wird (vgl. Borghoff u. a. 2003, S. 18). Jeff Rothenberg, ein amerikanischer Informationstechniker, hat innerhalb des Projekts NEDLIB eine Studie durchgeführt um zu testen, ob die Emulation als eine geeignete Strategie für die Langzeitarchivierung in Frage kommt. In seinem Bericht *An Experiment in Using Emulation to Preserve Digital Publications* schrieb er abschließend:

"The results of this study suggest that using software emulation to reproduce the behavior of obsolete computing platforms on newer platforms offers a way of running a digital document's original software in the far future, thereby recreating the content, behavior, and "look-and-feel" of the original document. The results of the initial iteration of the experiment described above indicate that this approach should work in principle, assuming that suitable emulators for obsolete computing platforms can be hosted on future platforms."

(Rothenberg 2000, S.83)

Jeff Rothenberg befand die Emulation als ein erfolgreiches Verfahren, deshalb wurden auch weitere Untersuchungen vorgeschlagen, aber anscheinend wurden sie noch nicht durchgeführt (vgl. Borghoff u. a. 2003, S. 259).

Die Vorteile der Emulation gegenüber der Migration sind:

- die Authentizität der Dokumente kann gewährleistet werden
- durch den Einsatz von Emulatoren können alle Dokumententypen dargestellt werden
- zu den digitalen Dokumenten bleiben auch die Programmstrukturen für die nachfolgenden Generationen erhalten
- der Aufwand der Emulation korreliert nicht mit dem Datenbestand

(vgl. Borghoff u. a. 2003, S. 81f.)

Diesen Vorteilen gegenüber steht die Frage:

Will der Nutzer, der z. B. in 30 Jahren, ein solches Dokument einsehen will überhaupt ein authentisches Dokument? Der informationelle Mehrwert der Dokumente wird bis dahin steigen und wie kann er dann ein so altes Dokument weiter bearbeiten? Auch die Peripherie wird bis dahin großen Änderungen unterliegen, der Trend geht hier zu Touch-Screen und Spracherkennung, kann der Nutzer dann mit unseren heutigen Gegebenheiten noch umgehen (vgl. Bischoff 2003)?

Logisch betrachtet scheint die Emulation die günstigere Lösung zu sein. Der Aufwand, der mit ihr einher geht, scheint geringer zu sein als der, der Migration. Außerdem ist es durch den Einsatz von Emulatoren möglich das Originaldokument zu betrachten, dies entspricht den *Kriterien für vertrauenswürdige Archive*, in denen die Authentizität der Dokumente gefordert wird. Andererseits fehlt bei der Emulation¹⁸ noch weitgehend die praktische Erfahrung und somit ist es schwierig die Risiken, die diese Strategie mit sich bringen könnte, einzuschätzen. Die Migration als Strategie wird so gut wie noch nicht umgesetzt, es fehlen noch

¹⁸ Erste Expertisen und Erfahrungen werden in einem Projekt am Lehrstuhl für Kommunikationssysteme an der Universität Freiburg gesammelt, weitere Informationen sind auf der Homepage unter <http://www.ks.uni-freiburg.de/fla/> zu finden.

Konzepte und Richtlinien. Deshalb gestaltet es sich schwierig das Für und Wider beider Strategien gegeneinander abzuwägen, wie auch, wenn keine genauen Kenntnisse über die Folgen vorliegen. Einzig klar ist, dass zu dieser Problematik schnellstmöglich eine Lösung gefunden werden muss, da die Übernahme von digitalen Dokumenten in digitale Archive, schon jetzt stattfindet (vgl. Bischoff 2003).

7 Rechtliche Aspekte

Die Deutsche Bibliothek sammelt unterschiedliche Arten von digitalen Publikationen. Es gibt abgabepflichtige elektronische Publikationen, die sie in ihrem gesetzlichen Auftrag sammelt, und solche, die sie zurzeit freiwillig sammelt, um ihrem Auftrag der lückenlosen Archivierung nachzukommen und um erste praktische Erfahrungen zu sammeln (vgl. Schwens 2002).

Hierzu ein Überblick:

Sammelpflichtige bzw. abgabepflichtige digitale Publikationen:

- Digitale Publikationen, die ein Verleger, gewerblich oder privat, veröffentlicht im Gebiet der Deutschen Bundesrepublik, Voraussetzung ist im Moment noch, dass diese auch auf einem physischen Datenträger gespeichert sind (vgl. DDB 2004e, S.1)

Die DDB sammelt freiwillig:

- Netzpublikationen von Kooperationspartnern (Verleger und verlegende Stellen), die sich für eine freiwillige Abgabe entschieden haben (vgl. DDB 2001)
- Netzpublikationen, deren Inhalt einmalig ist und welche nicht in anderer Form vorliegen z. B. Papier, Datenträger u. a. (vgl. Schwens 2002)
- Netzpublikationen, welche über öffentliche Kommunikationsnetze abrufbar sind (vgl. Schwens 2002)

Ein Teil der Netzpublikationen, die Die Deutsche Bibliothek freiwillig sammelt, sollen in Zukunft sammelpflichtig sein, die *Sammelrichtlinien und Auswahlkriterien*

hierfür werden vom *Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen für Deutschland* ausgearbeitet. Zusätzlich bedarf es einer Gesetzesnovellierung des *Gesetzes über die Deutsche Bibliothek (DBibIG)*, in der der Sammelauftrag verankert sein soll (Goebel u. a. 2004).

Des Weiteren werden auf dem Gebiet des Urheberrechts Gesetzesnovellierungen angestrebt, da der Zugriff für die Nutzer größtenteils nur über die Lesesäle Der Deutschen Bibliothek möglich ist und durch die Vergabe von Lizenzen, seitens der Verlage, hohe Kosten entstehen.

Hier die wichtigsten Vorschläge für Gesetzesnovellierungen des Urheberrechts:

- Bezug nehmend auf das Urheberrecht besteht der Wunsch Vervielfältigungen zu gestatten: bisher ist die Vervielfältigung nur für den privaten Gebrauch und sonstige eigene Zwecke gestattet, dies soll erweitert werden um die Vervielfältigungen von Publikationen aus Sammlungen, die einer nationalbibliographischen Aufgabe nachkommen oder einen ähnlichen kulturellen Zweck erfüllen, zu gestatten
- und es wird gefordert, dass die Vervielfältigungsstücke aus solchen Sammlungen öffentlich zugänglich gemacht werden dürfen solange keine kommerziellen Absichten dahinter stehen, also z. B. für Unterricht und Forschung
(Goebel u. a. 2004).

Die IST-Situation, wie sie momentan in Der Deutschen Bibliothek aussieht:

Der kostenlose Zugriff auf urheberrechtlich geschützte Dokumente, die auf dem Archivserver der DDB liegen, ist nur für Mitarbeiter der DDB gestattet und für registrierte Nutzer über die Multimedia-Arbeitsplätze in den Lesesälen der DDB (Frankfurt am Main/Leipzig) möglich (vgl. Schwens 2002). Der Download und Ausdruck von Teilen der Netzpublikation, die auf dem Archivserver liegt, kann an einer zentralen Stelle der DDB vorgenommen werden und ist kostenpflichtig (vgl. DDB 2001, S. 4).

Ein Zugriff von außerhalb kann nur auf copyrightfreie Publikationen gestattet

werden wie z. B. Online-Dissertationen, vorausgesetzt der Autor hat sein Einverständnis gegeben (vgl. Schwens 2002).

Mitarbeitern der DDB ist zusätzlich ein kostenloser Direktzugriff auf die Netzpublikationen der Verlage im Rahmen ihrer Archivierungs- und Erschließungstätigkeit gestattet (vgl. DDB 2001, S.4), ob dies dem Nutzer auch gestattet ist, ist abhängig von der jeweiligen Vereinbarung der DDB mit dem Anbieter.

Aus diesem Überblick ist ersichtlich, dass die Verwaltung der Zugriffsrechte sehr streng geregelt ist und sich recht unterschiedlich gestalten kann. Solange der Zugriff nur über die Nutzung der DDB erfolgen kann und die Vervielfältigung der Publikationen auf Grund des Urheberrechts nicht gestattet ist, scheint es für die Bibliotheksbenutzer im Allgemeinen kein attraktives Angebot zu sein und für die DDB andererseits ein großer Aufwand.

Um die Verwaltung der Rechte optimal gestalten zu können wünscht man sich von Seiten der Bibliothek noch ein Digital-Rights-Management (DRM). Dies ist ein System mit dessen Hilfe man die Bezugsrechte und –modalitäten der digitalen Dokumente verwalten kann. Im Rahmen des Projekts TECUP (Testbed implementation of the European Copyright User Platform)¹⁹, das vom 29.12.1998 bis zum 28.02.2001 lief, wurden Untersuchungen durchgeführt, welchen Anforderungen solch ein System Stand halten sollte (vgl. Dobratz u. a. 2001).

Soweit das für mich ersichtlich ist wurde hier die optimale Lösung noch nicht gefunden.

Über die DDB hinaus sollen die Regionalbibliotheken, für die Die Deutsche Bibliothek den Sammelauftrag ausführt, ein *regionales Fenster* erhalten, der den Zugriff auf die entsprechenden Dokumente zulässt, die in ihren Bereich fallen. Der Zugriff auf die elektronischen Publikationen ist analog zu dem der DDB geregelt (vgl. DDB 2001, S. 4).

¹⁹ <http://www.sub.uni-goettingen.de/gdz/tecup>

8 Projekte

In den Kapiteln 4 bis 7 wurden die organisatorischen Konzepte und weitere wesentlichen Aspekte zur Thematik der Langzeitarchivierung dargelegt. In diesem Kapitel möchte ich an Hand der Projekte einen Bezug zu Langzeitarchivierung in der Praxis herstellen. Diese vier Projekte, die hier kurz in ihren Grundzügen vorgestellt werden, sind in ihren Ansatzpunkten ganz verschieden und bieten einen Überblick über die Aktivitäten, die bezüglich der Langzeitarchivierung unternommen werden.

8.1 NESTOR

NESTOR steht für *Network of Expertise in long-term STOrage of digital Resources* und präsentiert das *Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen für Deutschland*. Die Laufzeit ist von Mitte 2003 bis Mitte 2006 angesetzt. Getragen wird das Projekt vom Bundesministerium für Bildung und Forschung (BMBF). Es ist ein Teilprojekt des *Kompetenznetzwerkes Neue Dienste, Standardisierung, Metadaten*, das auch vom BMBF getragen wird. Vorgänger war das Projekt *dl-konzepte* (Digital Library-Konzepte) aus dem die *Initiativegruppe Langzeitarchivierung* hervorgegangen ist, die wiederum in einem halbjährigen Folgeprojekt das Arbeitspapier *Langzeitverfügbarkeit digitaler Dokumente – Erarbeitung eines ersten kooperativen Konzeptes für Deutschland* erarbeitet hat (Workshop 2002). Dies war die "Initialzündung" um eine kooperative Langzeitarchivierungsstrategie und –policy in Deutschland zu entwickeln (vgl. Tiedau 2003, S. 1f.).

NESTOR setzt sich aus folgenden Kooperationspartnern zusammen: Die Deutsche Bibliothek, die Niedersächsische Staats- und Universitätsbibliothek Göttingen, die Bayrische Staatsbibliothek, der Computer und Medienservice der Universitätsbibliothek der Humboldt-Universität zu Berlin, dem Institut für Museumskunde in Berlin und der Generaldirektion der Staatlichen Archive Bayerns.

Die Kooperationspartner werden in ihrer Arbeit durch einen Lenkungsausschuss, dem Projektbeirat, begleitet, dieser setzt sich aus verschiedenen Fachvertretern zusammen. Sie kommen aus dem Bereich des Bibliothekswesens, des Archivwesens, der Wissenschaft, des Museumswesens, der Universitäten, des Verlagswesens, der Forschungsinstitute, der Rechenzentren, der Open Access Initiative oder sind einfach wissenschaftliche Nutzer (vgl. Schwens 2003, S. 3f.).

Durch diese Zusammensetzung der Kooperationspartner und des Projektbeirats sind alle Beteiligten, die sich mit der Archivierung und dem Veröffentlichen von digitalen Dokumenten befassen, mit einbezogen. Es ist nicht nur auf das Bibliothekswesen eingeschränkt, sondern bezieht auch die anderen Gedächtnisinstitutionen, Archive und Museen, mit ein (vgl. Tiedau, S. 2).

Langfristiges Ziel von NESTOR ist es eine dauerhafte, verteilte Infrastruktur für die Langzeitarchivierung und Langzeitverfügbarkeit digitaler Dokumente in Deutschland zu schaffen. Dies möchte man über eine Mitgliederorganisation lösen, Vorbild hierfür ist Großbritannien, dort wurde die Mitgliederorganisation unter dem Namen *Digital Preservation Coalition (DPC)*²⁰ gegründet (vgl. Tiedau, S.3).

Durch das Erreichen folgender Ergebnisse möchte man sich diesem Ziel in kleineren Schritten nähern (vgl. Tiedau, S. 3):

- (1) Aufbau eines webbasierten Forums²¹, das gleichzeitig als Informations- und Kommunikationsplattform für das Projekt genutzt wird; in diesem Forum werden alle inhaltlichen Angebote zum Thema Langzeitarchivierung und Langzeitverfügbarkeit digitaler Dokumente innerhalb Deutschlands bereitgestellt und es ist eine Expertendatenbank geplant, über die man kompetente Ansprechpartner finden kann
- (2) diese Plattform erhält ihre Informationen aus einer Informationsdatenbank
- (3) ein Kriterienkatalog für vertrauenswürdige Archive soll erarbeitet werden, dieser stützt sich auf die Ergebnisse, die RLG und OCLC in ihrem Papier *Trusted Digital Repositories: Attributes and Repositories* (RLG/OCLC 2002) zusammengefasst haben

²⁰ <http://www.dpconline.org>

²¹ <http://www.langzeitarchivierung.de>

- (4) um die Vertrauenswürdigkeit von Archivservern gewährleisten zu können sollen Empfehlungen für das Zertifizierungsverfahren ausgearbeitet werden
- (5) es werden Empfehlungen für Sammelrichtlinien und Auswahlkriterien für die digitalen Dokumente, die es zu archivieren gilt, erarbeitet
- (6) es soll eine "Policy" für die Langzeitarchivierung digitaler Dokumente konzipiert werden
- (7) um eine dauerhafte Organisationsform des Kompetenznetzwerkes und der dazugehörigen Plattform gewährleisten zu können soll ein Konzept entwickelt werden
- (8) und es soll ein detaillierter Arbeitsplan entworfen werden, in dem die verteilte Verantwortung und der Arbeitsablauf der beteiligten Institutionen geregelt ist

Die NESTOR-Plattform:

The screenshot shows the NESTOR website interface. At the top, there's a header with the NESTOR logo and the text 'Kompetenznetzwerk Langzeitarchivierung'. Navigation links for 'Login', 'Logout', 'english', and 'deutsch' are visible. The main content area is organized into several columns. On the left, there's a sidebar with 'Angebote' (Startseite, Projektinformationen, etc.) and 'Funktionen' (Suchen, Weiterempfehlen, etc.). The central part features a large article titled 'KOPAL - Kooperativer Aufbau eines Langzeitarchivs Digitaler Informationen' with a date of 'Am 1. Juli 2004'. To the right, there's a 'Termine' section listing events like 'Sun Summit Bibliotheken 2004' and 'Permanent Access to the Records of Science'. Below that is a 'Newsletter' subscription form with fields for 'Name' and 'Email', and a 'Go' button. At the bottom right, there's a 'Projektpartner' section listing institutions like 'Die Deutsche Bibliothek' and 'Bayer Staatsbibliothek'.

Abbildung 8: Nestor (Kompetenznetzwerk Langzeitarchivierung 2004)

Wie an Hand dieses Projekts ersichtlich ist steckt das Vorhaben eine Langzeitarchivierungsstrategie für Deutschland zu entwickeln noch in den Kinderschuhen. Die Infrastruktur für die Langzeitarchivierung digitaler Dokumente aus dem WWW muss erst noch aufgebaut werden.

8.2 Dissertationen Online

Das Projekt Dissertationen Online ging aus der IUK-Initiative hervor. Es begann im Januar 1998 und endete mit einer Laufzeit von über zwei Jahren im Oktober 2000. Ziel war die Entwicklung eines bundesweiten Konzepts für den Umgang mit digitalen Dissertationen, das von der Erstellung über das Retrieval bis hin zur Archivierung der Dissertationen reicht. Nach der Erprobung und der gegebenenfalls notwendigen Weiterentwicklungen soll das Konzept anschließend auf weitere Fachgebiete ausgedehnt werden. Dieses Ziel wurde erreicht, indem man das Vorhaben in kleinere Teilprojekte mit diversen Schwerpunkten untergliederte (vgl. dissonline 2000c):

(1) Metadaten

Hierzu wurde MetaDiss entwickelt, auf der Basis von DC, unter Einsatz von HTML 4.0. MetaDiss wurde inzwischen weiterentwickelt und seit September 2004 ist XMetaDiss im Einsatz, ebenfalls basierend auf DC, HTML wurde durch eine XML-DTD ersetzt, so ist es möglich eine klare Strukturierung der Inhalte durch Hierarchien zu gewähren (vgl. Korb 2004).

Der Datenaustausch erfolgt über eine OAI-Schnittstelle, weiterhin wird OAI-PMH (Open Archives Initiative – Protocol for Metadata Harvesting) eingesetzt um die Metadaten zu *harvesten*, *harvesten* bedeutet wörtlich übersetzt *ernten*. Dieses Verfahren erfolgt mittels eines Dataproviders, dies ist ein Datenanbieter mit einem Repository (Verwahrungsort) für Metadaten, und eines Serviceproviders, der die Daten des Dataproviders einsammelt bzw. als Harvester fungiert (vgl. Steinke 2004).

Zusätzlich wurden die Metadaten um technische Metadaten erweitert, speziell für die Langzeitarchivierung (vgl. Korb 2004).

Tools für die Erstellung von Metadaten sind im Informationssystem zu finden.

(2) Promotionsverfahren und Retrieval

Die Promotionsordnungen der Hochschulen sollen angeglichen werden und um die Abgabe der Dissertation in Form einer elektronischen Publikation erweitert werden. Von zahlreichen Hochschulen wurde dies schon umgesetzt (vgl. Wollschläger 2003).

Das Institute for Science Networking Oldenburg GmbH (ISN) hat sich mit der Thematik des Retrievals befasst. Ein wichtiges Aufgabengebiet des ISN ist es Bibliotheken mit verteilten Systemen zu unterstützen, es erstellt Konzepte und entsprechende Dienste für das Management von Informationen im Bereich der Wissenschaften, sowie Dokumenten- und Workflow-Systeme (vgl. ISN 2004).

(3) Formate

Es wurden Untersuchungen vorgenommen um festzustellen welche Formate sich am besten für die Archivierung und Präsentation eignen, sie wurden auch auf ihre Funktionalität bezüglich Konvertierung und Retrieval untersucht.

Prinzipiell nimmt Die Deutsche Bibliothek alle gängigen Formate an. Um die Langzeitverfügbarkeit der Dokumente gewährleisten zu können, bevorzugt sie bestimmte Dateiformate, diese sind hier in entsprechender Reihenfolge aufgelistet:

1. XML/SGML
2. HTML
3. PDF
4. PS
5. Sonstige (RTF, DOC, TEX, DVI, etc.)

(vgl. dissonline 2000b).

(4) Multimedia

Digitale Dissertationen weisen gegenüber gedruckten Dissertationen einen informationellen Mehrwert auf, der gleichbedeutend mit einem wissenschaftlichen Mehrwert ist. Deshalb wurden die Voraussetzungen für die Nutzung von

Strukturdaten untersucht, also in welcher Form die Rohdaten vorliegen sollen und wie sie aufbereitet werden sollen. Daraufhin wurden Bearbeitungsmethoden für entsprechende Multimedia-Formate entwickelt. Die multimedialen Hilfen und weitere Informationen stehen über das Informationssystem von Dissonline zur Verfügung (vgl. Diepold 2000).

(5) Beratung und Unterstützung

Es wurden Richtlinien und Empfehlungen für alle Beteiligten erarbeitet, die über das Informationssystem von Dissonline zur Verfügung stehen (vgl. Kamke 2000).

(6) Bibliotheksbereich

Für die beteiligten Bibliotheken wurde ein Workflow für den Geschäftsgang ausgearbeitet, sowie Informationen zu allen weiteren relevanten Themen ausgearbeitet, die auf der Homepage von Dissonline über den Button *Bibliotheken* nachgelesen werden können (vgl. dissonline 2000a).

(7) DDB

Das Projekt wurde in Zusammenarbeit mit Der Deutschen Bibliothek durchgeführt, ihre Aufgabe war es z. B. die Statistik der archivierten Online-Hochschulschriften zu führen oder die Organisation und Durchführung von Workshops und Tagungen u. a. zu übernehmen (vgl. Kamke 2000). Des Weiteren fungiert Die Deutsche Bibliothek anschließend als Koordinierungsstelle (vgl. Wollschläger 2003).

Betreut wurden die Teilprojekte durch den Bereich *Projektkoordination- und Management*, das ein eigenes Teilprojekt darstellte (vgl. Kamke 2000). Parallel zum Projekt wurde das Informationssystem von Dissonline aufgebaut, als Wegweiser für alle, die sich mit dieser Thematik befassen, dies verbirgt sich hinter dem Namen *FAQ—Fragen und Antworten rund um die Online-Dissertationen*.

Zurzeit nehmen 82 Hochschulen am Ablieferungsverfahren von Online-Hochschulschriften teil. Der Lieferweg für die Hochschulschriften erfolgt über das Push-Verfahren²², via E-mail oder als FTP (File Transfer Protocol)-Upload. Ziel ist es die Online-Hochschulschriften über einen automatisierten Pull-Dienst an Die Deutsche Bibliothek abzuliefern (vgl. Steinke 2004).

²² *Push* kommt aus dem Englischen und wird hier im Sinne von *Schieben* verstanden, der Gegensatz hierzu ist das Pull-Verfahren, hierunter versteht man das *Ziehen* von Dateien

Die Anzahl der archivierten Online-Hochschulschriften ist seit Ende 1998 drastisch gestiegen:

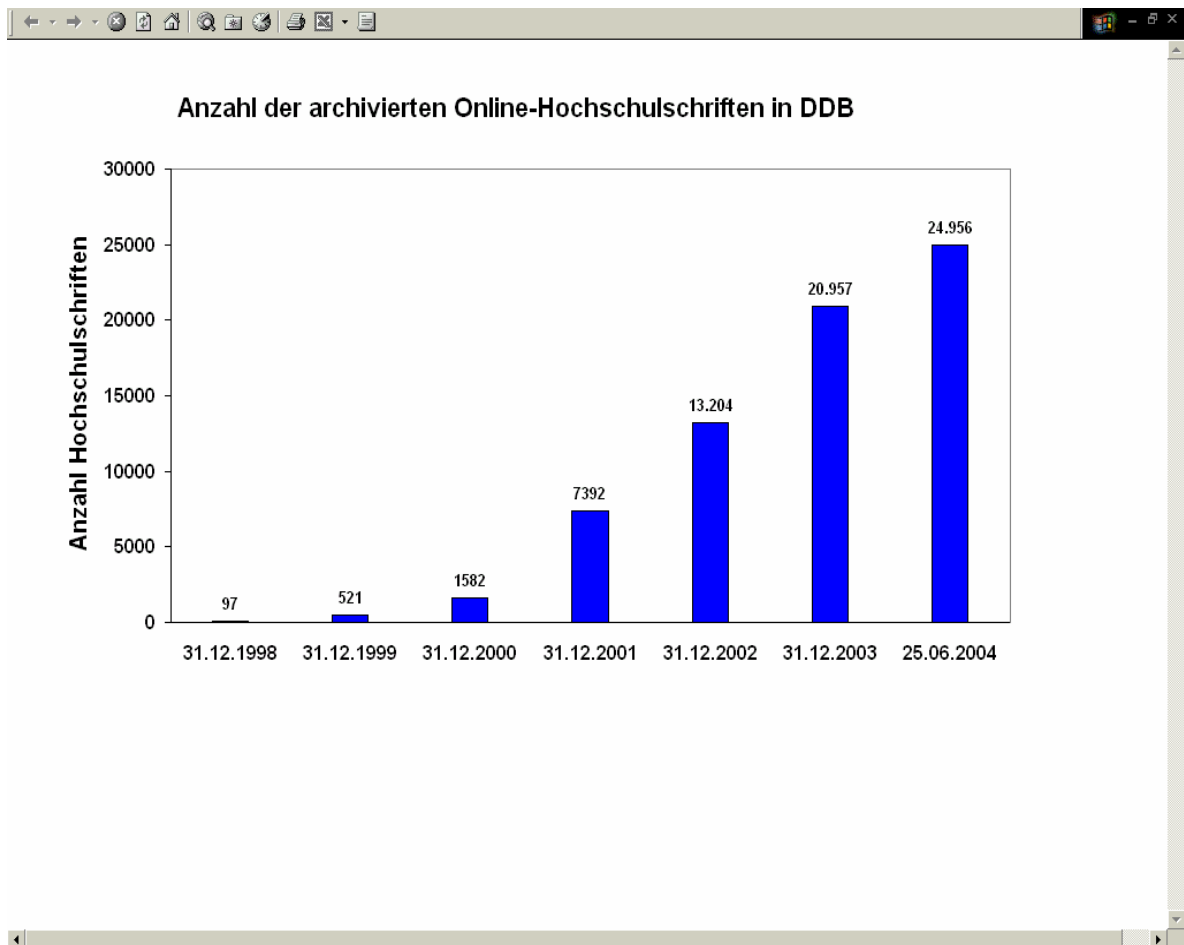


Abbildung 9: Statistik Online-Hochschulschriften (DDB 2004f)

Im Anschluss an das Projekt Dissertationen Online wurde zum 01.02.2001 an Der Deutschen Bibliothek die *Koordinierungsstelle Dissonline* ins Leben gerufen. Ihre Aufgabe ist es alle beteiligten Institutionen zu beraten und für die Koordinierung zu sorgen, sowie die Ergebnisse von Dissertationen Online weiter zu entwickeln.

Zum 01.01.2003 folgte da Projekt *Aufbau einer Koordinierungsstelle für elektronische Hochschulschriften*, das den Ausbau der *Koordinierungsstelle Dissonline* als Ziel hat. Über dieses Projekt erfolgte u. a. eine Analyse der Infrastruktur bezüglich Online-Dissertationen, sowie Vergleiche der

Promotionsordnungen, Geschäftsgänge und Online-Angebote der beteiligten Institutionen. Die Hochschulen wurden zusätzlich mittels einer Umfrage um Anregungen gebeten. Auf Grund der daraus resultierenden Ergebnisse wurde ein interaktives Informationssystem aufgebaut. Der Einstieg kann nun über die Zielgruppe (Bibliotheken, Doktoranden, Fachbereiche, Technik) erfolgen oder thematisch (Veröffentlichen, Recherche, Rechtsfragen, Formate, Bibliothek, Doktorand, Fachbereich, Technik) gewählt werden (vgl. Wollschläger 2003).

Die Website von Dissonline



Das Projekt Dissertationen Online war ein sehr umfangreiches Projekt, das ich hier nur oberflächlich anschneiden konnte.

8.3 BOA (Baden-Württembergisches Online-Archiv)

Im Frühjahr 2003 begannen die Landesbibliotheken von Baden-Württemberg (Karlsruhe und Stuttgart) in Kooperation mit dem BSZ in Konstanz sich mit der Langzeitarchivierung von Online-Publikationen zu befassen. Im Herbst 2003 erhielt das Projekt den Namen BOA (Baden-Württembergisches Online-Archiv). Ziel des Projekts war es gemeinsam eine technische Plattform zu entwickeln, sowie einen Workflow für den Geschäftsgang von Netzpublikationen zu erarbeiten.

Da bis zum heutigen Zeitpunkt noch keine Sammelrichtlinien und Auswahlkriterien existieren, diese sind noch in Arbeit, sammeln die Regionalbibliotheken Netzpublikationen, die einen Pflichtexemplar-Charakter aufweisen oder solche, die von landeskundlicher Relevanz sind, sowie jene, die für die Sondersammlungen anderer beteiligten Bibliotheken von Relevanz sind. Vorerst kommt hinzu, dass die Netzpublikationen im Internet frei verfügbar sind, da die rechtlichen Rahmenbedingungen für andere Internetressourcen noch nicht geschaffen sind.

Die Dokumente werden nach RAK-NBM im Verbundkatalog bzw. in der Zeitschriftendatenbank (ZDB) erfasst und nach RSWK verschlagwortet. Das Katalogisat wird nach MAB2 (Maschinelles Austauschformat für Bibliotheken Version 2) transferiert, deshalb liegen die bibliographischen Daten im MAB-Format vor, als Ausgabeformat soll jedoch auch DC verwendet werden. Weiterhin werden Dewey Decimal Classification-Sachgruppen vergeben und Notationen der Landesbibliographie von Baden-Württemberg.

Als Software hatte sich die Software des Medienservers OPUS (Online Publikationsverbund Stuttgarts) angeboten, da die OPUS-Software im Stuttgarter Raum schon für die Bereitstellung von Hochschulschriften auf Medienservern im Einsatz ist. Da OPUS aber nicht für hierarchisch verknüpfte Elemente geeignet ist, ist man inzwischen auf die JAVA-Anwendung ESEM umgestiegen, die vom BSZ entwickelt worden war. Mit Hilfe dieser Software ist es jetzt möglich verschiedene Hierarchieebenen anzulegen. Hierarchieebenen werden für zeitschriftenartige Publikationen und für Websites benötigt.






Nachdem das Dokument erschlossen ist, muss nur noch die entsprechende URL eingegeben werden. Um die Langzeitarchivierung gewährleisten zu können werden die Netzpublikationen über eine URN bei Der Deutschen Bibliothek registriert. Gespeichert werden die Dokumente auf dem Archivserver des BSZ. Die Dokumente liegen im PDF-Format vor oder als HTML-Dokument. Letztere werden mittels des Offline-Browser HTTRACK abgespeichert. Dieser arbeitet sich Link für Link durch das Dokument und speichert jede einzelne Seite ab.

Noch ist es aus technischen Gründen nicht möglich Web-Datenbanken und dynamische Websites zu archivieren.

Mit diesem Projekt wurde nur der erste Schritt in Richtung Langzeitarchivierung unternommen. Die Dokumente sind in BOA gespeichert und für den Benutzer zugänglich, zusätzlich muss jedoch auch die Langzeitverfügbarkeit gewährleistet werden können, die nur über den Einsatz von Migration und Emulation erreicht werden kann. Da die Konzepte für die Umsetzung dieser Strategien noch nicht ausgereift sind, werden vorab technische Metadaten mit erfasst. Hierzu zählen Dateigröße, Format, Erstellungssoftware (wenn möglich) und Präsentationssoftware. Die Erstellungssoftware ist ausschlaggebend für die Migration, die Präsentationssoftware für die Emulation. So ist es möglich beiden Strategien gegenüber offen zu bleiben. Der Benutzer wird nach den Dokumenten im lokalen OPAC, im Südwestverbund (SWB) oder im BOA-System recherchieren können. Per Mausklick gelangt er direkt zu Dokument.

Noch ist BOA im Aufbau und die Anzahl der enthaltenen Dokumente sehr gering, sobald eine "kritische Masse" verfügbar gemacht wurde, soll BOA der breiten Öffentlichkeit vorgestellt werden (vgl. Wiesenmüller 2004).

Die BOA-Website:

	<div style="display: flex; justify-content: space-around; align-items: center;">  <div style="text-align: center;"> BADISCHE LANDESBIBLIOTHEK </div>  <div style="text-align: center;"> Württembergische Landesbibliothek </div> </div>
<div style="background-color: #ffffcc; padding: 5px;"> Home erweiterte Suche Browsen Volltextsuche Editieren Hilfe </div>	<div style="text-align: center; border-bottom: 1px solid black; padding-bottom: 5px;"> Baden-Württembergisches Online-Archiv (BOA) </div> <div style="margin-top: 10px;"> Recherche: <input style="width: 200px;" type="text"/> <input type="button" value="Suche starten"/> </div> <p>Veröffentlichungen, die in Baden-Württemberg erscheinen oder sich inhaltlich mit dem Land, seinen Orten und Personen beschäftigen, werden an der Badischen Landesbibliothek in Karlsruhe und der Württembergischen Landesbibliothek in Stuttgart gesammelt, erschlossen, der Öffentlichkeit zur Verfügung gestellt und für die Zukunft archiviert. Waren dies früher überwiegend gedruckte Materialien, so werden künftig auch relevante Internet-Angebote einbezogen. Gemeinsam mit dem Bibliotheksservice-Zentrum Baden-Württemberg wird deshalb derzeit das "Baden-Württembergische Online-Archiv (BOA)" aufgebaut.</p> <p>BOA befindet sich noch in der Entwicklungsphase - bitte haben Sie Verständnis, dass gelegentlich Fehler auftauchen und noch nicht alle Funktionen eingebaut sind.</p> <p>Wenn Sie mehr über das Projekt BOA wissen wollen: Im 'Bibliotheksdienst' Jg. 38 (2004), Heft 4, S. 471-479, erschien ein Artikel dazu, den wir mit freundlicher Genehmigung der Redaktion online zur Verfügung stellen:</p> <p>Wiesenmüller, Heidrun: Langzeitarchivierung von Online-Publikationen an Regionalbibliotheken: Das Projekt 'Baden-Württembergisches Online-Archiv' (BOA)</p> <p>Ansprechpartner:</p> <ul style="list-style-type: none"> • BLB: Christoph Höck • WLB: Heidrun Wiesenmüller • BSZ/Technik: Andreas Schack
<div style="background-color: #ffffcc; padding: 10px; text-align: center;">   </div>	

8.4 SpringerLink

SpringerLink ist ein internetbasierter Informationsdienst des Springer-Verlags. Mittels SpringerLink erhält man Zugriff auf ca. 500 Zeitschriften, 1.600 Büchern, zwei Lernsoftwares und 5 Expertensysteme aus elf verschiedenen Fachgebieten, vorwiegend aus dem STM (Science Technology Medicine)-Bereich. Es können durchschnittlich 5 Millionen Zugriffe im Monat verzeichnet werden (vgl. Springer-Verlag und Die Deutsche Bibliothek kooperieren in der Langzeitarchivierung von Online-Publikationen 2002).

Im Jahre 1998 hat der Springer-Verlag in Kooperation mit Der Deutschen Bibliothek ein Pilotprojekt begonnen. Ziel war es die Zeitschriften und

Monographien, die über SpringerLink angeboten werden, langfristig zu archivieren. Die Deutsche Bibliothek hat diese Kooperation sehr begrüßt, einerseits, um ihrem Auftrag der lückenlosen Archivierung aller deutschsprachigen und deutschen Veröffentlichungen nachkommen zu können, andererseits, um erste Erfahrungen in der Praxis sammeln zu können (vgl. Springer-Verlag und Die Deutsche Bibliothek kooperieren in der Langzeitarchivierung von Online-Publikationen 2002).

Bis Dezember 2003 konnten 418 elektronische Zeitschriften und 1.300 Monographien archiviert werden. Sie wurden katalogisiert und anschließend auf dem Archivserver der DDB archiviert. Nachgelieferte Metadaten des Springer-Verlags wurden nicht berücksichtigt, da von den vorliegenden Printpublikationen schon Katalogisate vorhanden waren, die als optimaler Ausgangspunkt angesehen wurden. Die Archivierung der digitalen Publikationen erfolgt mittels Spiegelung von SpringerLink auf den Archivserver Der Deutschen Bibliothek. Es konnte ein Datenlieferungsverfahren (LinkDelivery Service) und Speicherstrukturen für fortlaufende elektronische Publikationen entwickelt werden (vgl. Liegmann 2003).

Nachdem die DDB auf Grund dieses Pilotprojekts erste positive Erfahrungen gesammelt hatte, wurde im Frühjahr 2002 die *Rahmenvereinbarung zur freiwilligen Ablieferung von Netzpublikationen zum Zwecke der Verzeichnung und Archivierung* (DDB 2001) zwischen Der Deutschen Bibliothek und dem Börsenverein des Deutschen Buchhandels abgeschlossen. Diese Rahmenvereinbarung dient der Überbrückung bis zur Gesetzesnovellierung, in der entsprechende Netzpublikationen einen Pflichtexemplar-Charakter erhalten sollen (vgl. Springer-Verlag und Die Deutsche Bibliothek kooperieren in der Langzeitarchivierung von Online-Publikationen 2002).

In kleinen Schritten soll das Verfahren der Langzeitarchivierung elektronischer Publikationen optimiert werden. Wünschenswert ist es eine Standardisierung des Übergabeverfahrens zu entwickeln, sowie ein elektronisches Datenmodell für digitale Zeitschriftenartikel, in dem Primärdaten und deskriptive Metadaten in einem Package (vgl. Kapitel 3, S. 12/13) zusammengefasst sind. Der Benutzerkomfort soll ausgebaut werden, da der Retrieval über Metadaten oder Volltext noch nicht möglich ist. Das Verfahren der Spiegelung ist nur eine

Übergangslösung, sie soll durch das Pärchen-Lieferprinzip ersetzt werden, also Lieferung von Primärdaten und Metadaten in einem Package (vgl. Liegmann 2003).

9 Schlussbetrachtung

Die Thematik der Langzeitarchivierung digitaler Dokumente aus dem WWW ist ein sehr komplexes Thema. Wie aus dieser Diplomarbeit ersichtlich ist, sind es die ersten Schritte, die diesbezüglich unternommen wurden.

Noch ist es nicht abzusehen, wie das OAIS-RM genau umgesetzt werden soll, wir sind hier am Anfang der Theorien, die anschließend erst in der Praxis erprobt werden müssen.

Ebenso ist es mit den Strategien der Langzeitarchivierung, Migration und Emulation. Es müssen vorab Konzepte erarbeitet und Erfahrungen in der Umsetzung gesammelt werden, um Folgen und Risiken abschätzen zu können.

Auf der Benutzerseite ist der Handlungsbedarf ebenfalls noch groß. Der Benutzerkomfort, bezüglich Retrieval und Darstellung, sollte weiter ausgebaut werden, sowie die gesetzlichen Regelungen der Zugriffsrechte vereinfacht werden. Ferner sollte der Zugriff kostengünstiger werden. Sonst steht der Aufwand der Langzeitarchivierung nicht in Korrelation mit den Kosten und der Nutzung.

Des Weiteren ist fraglich, ob der Nutzer die Dokumente im Original-Format betrachten will, so wie es in den *Kriterien für vertrauenswürdige Archive* gefordert wird, da man davon ausgehen kann, dass der Nutzer die Dokumente weiterverarbeiten will und die Original-Dokumente dem informationellen Mehrwert der Zukunft nicht Stand halten können.

Sicher ist, dass schnellstmöglich Lösungen gefunden werden müssen, da die Technik sich rasant entwickelt und die Anzahl der digitalen Dokumente weiter steigt. Die Bibliotheken müssen versuchen trotz mangelndem Budget mit der Entwicklung in der Informations-Technologie Schritt zu halten. Daher werden immer mehr Projekte ins Leben gerufen, die sich der Erprobung in der Praxis dieses Themas, stellen.

Literaturverzeichnis

akademie.de 2004 akademie.de: *Net-Lexikon*. [2004]. URL: <http://www.lexikon-definition.de/Net-Lexikon.html>. Zugriff am: 2004-07-15

Arbeitsgruppe Digitalisierung 1996 Arbeitsgruppe Digitalisierung: *Digitalisierung gefährdeter Bibliotheks- oder Archivguts*. 1996-10-07. URL: <http://www.lad-bw.de/lad/dfgdigh1.htm>. Zugriff am: 2004-04-10

Bischoff 2003 Bischoff, Frank M.: Migration, Emulation, Konversion: Möglichkeiten und Grenzen. In: *Archiv und Wirtschaft* Jg. 36 (2003), Nr. 3, S. 119 – 127

Börsenverein des Deutschen Buchhandels 1996 Börsenverein des Deutschen Buchhandels: *Die unendliche Bibliothek : digitale Information in Wissenschaft, Verlag und Bibliothek*. - Wiesbaden : Harrassowitz , 1996. - 120 S. - ISBN 3-447-03785-7

Borghoff u. a. 2003 Langzeitarchivierung : Methoden zur Erhaltung digitaler Dokumente / Uwe M. Borghoff ... - 1. Aufl. Heidelberg : dpunkt-Verl., 2003. - XV, 283 S. : Ill. - ISBN 3-89864-245-3

BSB 2000 Bayrische Staatsbibliothek: *Erschließung elektronischer Online-Ressourcen*. 2000-09-29. URL: http://www.bib.bvb.de/bib_schule/bib_sch6.htm. Zugriff am: 2004-08-07

Byrum 2002 Byrum, John D.: Herausforderungen bei der Verzeichnung von Netzpublikationen in Nationalbibliographien: Ein Überblick über Aufgaben und Lösungsansätze. In: *68th IFLA Council and General Conference August 18-24, 2002*. 2002-08-24. URL: <http://www.ifla.org/IV/ifla68/papers/117-152g.pdf>. Zugriff am: 2004-08-04

Carmen 2000 Content Analysis, Retrieval, Metadata: Effective Networking: *Carmen AP/2 Projektinhalt*. 2000-06-24. URL: <http://www.sub.uni-goettingen.de/carmen/Projektinhalt/projektinhalt.html>. Zugriff am: 2004-08-23

CCSDS 2002a Consultative Committee for Space Data Systems: *Reference model for an open archival information system (OAIS)*. 2002-01-01. URL:

<http://www.ccsds.org/docu/dscgi/ds.py/Get/File-143/650x0b1.pdf>. Zugriff am: 2004-04-25

CCSDS 2002b Consultative Committee for Space Data Systems: *The Open Archival Information System (OAIS) Reference Model and its Usage*. [2002]. URL:

http://www.ccsds.org/documents/so2002/spaceops02_p_t5_39.pdf. Zugriff am: 2004-04-25

DCMI 2003 Dublin Core Metadata Initiative: *Dublin Core Metadata Element Set, Version 1.1: Reference Description*. 2003-06-02. URL:

<http://dublincore.org/documents/dces/>. Zugriff am: 2004-08-11

DDB 2001 Die Deutsche Bibliothek: *Rahmenvereinbarung zur freiwilligen Ablieferung von Netzpublikationen zum Zwecke der Verzeichnung und Archivierung*. 2001-03-19. URL:

http://www.deposit.ddb.de/netzpub/web_rahmenvereinbarung.htm. Zugriff am: 2004-09-01

DDB 2002 Die Deutsche Bibliothek: *Die Deutsche Bibliothek 2002*. 2003-08-22. URL: [http://www.ddb.de/cgi-](http://www.ddb.de/cgi-bin/bermudix.pl?url=suche/suche.htm&men=suche/menu_txt.htm)

[bin/bermudix.pl?url=suche/suche.htm&men=suche/menu_txt.htm](http://www.ddb.de/cgi-bin/bermudix.pl?url=suche/suche.htm&men=suche/menu_txt.htm). Zugriff am: 2004-06-01

DDB 2003 Die Deutsche Bibliothek: *Standardisierungsarbeit für Bibliotheken : Konzept Der Deutschen Bibliothek*. 2003-05-13. URL:

<http://www.ddb.de/professionell/pdf/konzept.pdf>. Zugriff am: 2004-08-12

DDB 2004a Die Deutsche Bibliothek: *MAB*. 2003-12-08. URL:

<http://www.ddb.de/professionell/projekte.htm>. Zugriff am: 2004-08-21

DDB 2004b Die Deutsche Bibliothek: *Projekte*. 2004-07-15. URL:

<http://www.ddb.de/professionell/mab.htm>. Zugriff am: 2004-08-11

DDB 2004c Die Deutsche Bibliothek: *Regeln für die Erschließung elektronischer Ressourcen*. 2004-03-15. URL: http://www.ddb.de/professionell/afs_kern.htm.

Zugriff am: 2004-08-08

DDB 2004d Die Deutsche Bibliothek: *Sammelauftrag Der Deutschen Bibliothek*. 2004-07-20. URL: <http://www.ddb.de/wir/sammelauftr.htm>. Zugriff am: 2004-09-01

DDB 2004e Die Deutsche Bibliothek: *Sammlung, Verzeichnung und Archivierung von Netzpublikationen*. 2004-05-21. URL: <http://www.ddb.de/wir/netzpubl.htm>. Zugriff am: 2004-09-01

DDB 2004f Die Deutsche Bibliothek: *Online-Hochschulschriften : Sammlung, Verzeichnung und Archivierung durch Die Deutsche Bibliothek*. 2004-06-25. URL: http://www.ddb.de/netzpub/web_online-hochschulschriften_stat.htm.htm. Zugriff am: 2004-09-30

Diekmann u. a. 2002 *Langzeitverfügbarkeit digitaler Dissertationen – das Projekt CARMEN-AP4 „Persistent Identifier“ und die Koordinierungsstelle DissOnline* / Diekmann, Bernd... [ca. 2002-02]. URL: <http://www.bis.uni-oldenburg.de/~diekmann/t2.doc>. Zugriff am: 2004-04-10

Diepold 2000 Diepold, Peter: G. Ergebnisse und Ausblick. [2000-10]. (Das DFG-Projekt „Dissertationen Online“ stellt sich vor). In: *B.I.T.online* Jg. 3 (2000), Nr. 4, S. 422 - 428

dissonline 2000a Dissertationen Online: *Bibliothek und Rechenzentrum – Von der Notwendigkeit der Zusammenarbeit*. [2000]. URL: http://www.dissonline.de/texte_html/kolab.html. Zugriff am: 2004-10-03

dissonline 2000b Dissertationen Online: *Formate*. [2000]. URL: <http://www.deposit.ddb.de:8080/cocoon/xml-xsl/Dissonline/thema4-01.xml>. Zugriff am: 2004-10-03

dissonline 2000c Dissertationen Online: *Bibliothek und Rechenzentrum – Von der Notwendigkeit der Zusammenarbeit*. [2000]. URL: <http://www.dissonline.de/html/history.html>. Zugriff am: 2004-10-03

Dobratz u. a. 2001 Dobratz, Susanne ; Liegmann, Hans ; Tappenbeck, Inka: Langzeitarchivierung digitaler Dokumente. In: *Zeitschrift für Bibliothekswesen und Bibliographie* Jg. 48 (2001), Nr. 6, S. 327 – 332

Dobratz 2002 Dobratz Susanne: Produktion digitaler Publikationen. In: *Workshop „Langzeitverfügbarkeit digitaler Dokumente – Erarbeitung eines ersten kooperativen Konzepts für Deutschland“ : Arbeitspapier*. 2002.10-04. URL: <http://www.dl-forum.de/Foren/Langzeitverfuegbarkeit/Arbeitspapier2.pdf>. Zugriff am: 2004-08-10

Dobratz und Tappenbeck 2002 Dobratz, Susanne ; Tappenbeck, Inka: Thesen zur Zukunft der digitalen Langzeitarchivierung in Deutschland. In: *Bibliothek, Forschung und Praxis* Jg. 26 (2002), Nr. 3, S. 257 – 261

Goebel u. a. 2004 Goebel,... ; Scheller,... ; Zimmermann,...: *Digitale Langzeitarchivierung und Recht*. 2004-05. URL: http://www.isn-oldenburg.de/nestor-workshop/slides/slides_zimmermann.pdf. Zugriff am: 2004-10-04

Hashing 2002 *Hashing*. 2002. URL: http://www.dpunkt.de/java/Programmieren:mit_Java/Sicherheit/12.html. Zugriff am: 2004-08-13

Egger 2003 Egger, Alexander: *Digitale Signaturen, Probleme und Lösungen bei der Archivierung*. 2003-12-01. URL: <http://www.voeb.uibk.ac.at/odok2003/egger.pdf>. Zugriff am: 2004-08-13

Endres und Fellner 2000 Endres, Albert ; Fellner, Dieter W.: *Digitale Bibliotheken : Informatik-Lösungen für digitale Wissensmärkte*. - 1. Aufl. - Heidelberg : dpunkt-Verl., 2000. - 494 S.
ISBN 3-932588-77-0

Engster 2002 Engster, Florian: *Digitale Bibliothek : Konzeption und Implementierung mit der Greenstone Digital Library Software*.

Stuttgart, Fachhochschule Stuttgart – Hochschule der Medien, Diplomarbeit, 2002

EPICUR 2004a Enhancement of Persistent Identifier Services – Comprehensive Method for unequivocal Resource Identification: *EPICUR-Realisierungskonzept (Kurzbeschreibung)*. 2004-01-27. URL: <http://www.persistent-identifier.de/?link=336>. Zugriff am: 2004-06-03

EPICUR 2004b Enhancement of Persistent Identifier Services – Comprehensive Method for unequivocal Resource Identification: *Statistik : Uniform Resource Names (URN) für Netzpublikationen*. 2004-08-06. URL: <http://www.persistent-identifier.de/?link=542>. 2004-08-23

Extensible Markup Language 2004 akademie.de: Extensible Markup Language. In: *Net-Lexikon*. [2004]. URL: <http://www.lexikon-definition.de/Extensible-Markup-Language.html>. Zugriff am: 2004-07-15

Feather 2004 *Managing preservation for libraries and archives : current practice and future developments* / edited by John Feather. - Aldershot, Hants, England : Ashgate, 2004. - xi, 181 p.
ISBN 0-7546-0705-4

ISN 2004 Institute for Science Networking Oldenburg GmbH: *Profil*. 2004-06-08. URL: <http://www.isn-oldenburg.de/profil.html>. Zugriff am: 2004-10-03

Kamke 1999 Kamke, Hans-Ulrich: A. Überblick über das Projekt. [1999-10]. (Das DFG-Projekt „Dissertationen Online“ stellt sich vor). In: B.I.T.online Jg. 2 (1999), Nr. 4, S. 424 – 430

Kampffmeyer und Rogalla 1997 Kampffmeyer, Ulrich ; Jörg Rogalla: Grundsätze der elektronischen Archivierung : code of practice zur revisionssicheren Archivierung . - 2. überarb. u. erw. Aufl. - Darmstadt : VOI Verband Optische Informationssysteme e.V., 1997. - XII, 113 S.

(VOI-Schriftenreihe Kompendium ; 3)

ISBN: 3-932898-03-6 A

Kett, Jürgen: MABxml-1.2004-02-04. URL: http://www.ddb.de/professionell/pdf/mabxml_1_dok.pdf. Zugriff am: 2004-06-03

Kompetenznetzwerk Langzeitarchivierung 2004a Kompetenznetzwerk Langzeitarchivierung: *Glossar : Persistent Identifier*. 2004-02-20. URL: http://www.langzeitarchivierung.de/index.php?module=Encyclopedia&func=display_term&id=27&vid=1. Zugriff am 2004-08-23

Kompetenznetzwerk Langzeitarchivierung 2004b Kompetenznetzwerk

Langzeitarchivierung: *Langzeitarchivierung digitaler Ressourcen*. 2004-03-17.

URL: <http://www.langzeitarchivierung.de/modules.php?op=modload&name=News&file=article&sid=5&mode=thread&order=0&thold=0>. Zugriff am 2004-04-10

Kompetenznetzwerk Langzeitarchivierung 2004c Kompetenznetzwerk

Langzeitarchivierung: Langzeitarchivierung digitaler Dokumente in Deutschland - Projektbeschreibung. 2004-07.

URL: <http://www.langzeitarchivierung.de/modules.php?op=modload&name=News&file=article&sid=9>. Zugriff am: 2004-04-10

Korb 2003 Korb, Nikola: *Online-Dissertationen für die Praxis*. 2003-03-10. URL:

<http://www.iwi-iuk.org/iuk2003/program/korb/ppt/d.ppt>. Zugriff am: 2004-04-10

Liegmann 2001 Liegmann, Hans: Langzeitverfügbarkeit digitaler Publikationen.

[2001]. URL: <http://www.uni-muenster.de/Forum-Bestandserhaltung/konversion/digi-liegmann.shtml>. Zugriff am: 2004-04-15

Liegmann 2003 Liegmann, Hans: Langzeitarchivierung SpringerLink durch Die

Deutsche Bibliothek. In: *Dialog mit Bibliotheken* Jg. 15 (2003), Nr. 1, S. 26 – 28

Liegmann 2004a Liegmann, Hans: *Langzeiterhaltung digitaler Ressourcen -*

Aktivitäten & Kooperationen. 2004-02-17. URL:

http://www.langzeitarchivierung.de/downloads/Inetbib2003_Liegmann.ppt. Zugriff am: 2004-04-26

Liegmann 2004b Liegmann, Hans: *Langzeitverfügbarkeit elektronischer*

Publikationen Pilotprojekt des Springer-Verlages und Der Deutschen Bibliothek.

2004-05-17. URL: http://deposit.ddb.de/netzpub/web_koop_springer.htm. Zugriff am: 2004-06-03

Lupovici und Masanès 2000 Lupovici, Catherine ; Masanès, Julien: *Metadata*

for Long Term Preservation. 2000-07. URL:

<http://www.kb.nl/coop/nedlib/results/NEDLIBmetadata.pdf>. (NEDLIB Report series ; 2). Zugriff am: 2004-08-30

Lupprian 2004 Lupprian, Karl-Ernst: *Zur Langzeitarchivierung im Archivwesen*. 2004-03-25. URL:

http://www.langzeitarchivierung.de/downloads/Leipzig2004_Lupprian.ppt. Zugriff am: 2004-04-10

Malle 1996 Malle, Bernhard: Ein Beitrag zur Langzeitarchivierung von Produkten / Bernhard Malle. - Hamburg : Kovac, 1997. - XII, 228 S.; (dt.)

ISBN 3-86064-536-6

Zugl.: Karlsruhe, Univ., Diss., 1996

MP3 2004 akademie.de: MP3. In: *Net-Lexikon*. [2004]. URL:

<http://www.lexikon-definition.de/MP3.html>. Zugriff am: 2004-07-31

NEDLIB Consortium 2000 NEDLIB Consortium: *Standards for DSEP – Standards for the Implementation of a Deposit System for Electronic Publications (DSEP)*. 2000. URL: <http://www.kb.nl/coop/nedlib/results/dsepstandards.pdf>.

(NEDLIB Report series ; 4). Zugriff am: 2004-04-25

Neumann 2003 Neumann, Claudia: *Nachhaltige Nutzung digitaler Dokumente* / Claudia Neumann.

Stuttgart, Fachhochschule Stuttgart – Hochschule der Medien, Diplomarbeit, 2003

Niggemann, Elisabeth u. a. 2001 *Langzeiterhaltung digitaler Informationen – Networked European Deposit Library NEDLIB*. 2001. URL:

http://www.klostermann.de/zeitsc/zfbb_483.htm. Zugriff am: 2004-04-25

Oberhauser 2004 Oberhauser, Otto: Langzeitarchivierung als Thema internationaler Institutionen und Organisationen: Ein bibliotheksorientierter Überblick. In: *Zeitschrift für Bibliothekswesen und Bibliographie* Jg. 50 (2004), Nr.6, S. 303 – 314

OCLC 2002 Online Computer Library Center: *Web Characterization*. [2002]. URL: <http://wcp.oclc.org/>. Zugriff am: 2004-07-31

OCLC/RLG 2001 Online Computer Library Center und Research Libraries Group: *Preservation Metadata for Digital Objects : A Review of the State of the Art : A White Paper by the OCLC/RLG Working Group on Preservation Metadata*. 2001-01-31. URL: http://www.oclc.org/research/projects/pmwg/presmeta_wp.pdf. Zugriff am: 2004-08-23

ÖNB 2003 Österreichische Nationalbibliothek: *Grundlagen digitaler Archive*. 2003-09-20. URL: <http://onb.ac.at/about/lza/archiv.htm>. Zugriff am: 2004-06-03

Ostermann 1999 Ostermann, Raphael: Potentielle Dateiformate zur Langzeitarchivierung von Dokumenten unter Berücksichtigung von Primär- und Metainformationen. In: *Digitale Herausforderungen für Archive : 3. Tagung des Arbeitskreises "Archivierung von Unterlagen aus Digitalen Systemen" am 22. und 23. März 1999 im Bundesarchiv in Koblenz* / hrsg. von Michael Wettengel. - Koblenz, 1999. - 139 S. : Ill.

(Bundesarchiv <Koblenz>: Materialien aus dem Bundesarchiv ; 7)
ISBN 3-89192-086-5

Oßwald 2004 Oßwald, Achim: *Archivierungskonzepte für digitale Publikationen*. 2004-06-18. URL: http://www.fbi.fh-koeln.de/institut/personen/osswald/Material_Osswald/ss03/Archivierungskonzepte040403_3auf1.pdf. Zugriff am :2004-08-03

Payer 2002a Payer, Margarete: *Preservación del patrimonio cultural en la era de la digitalización : una tarea conjunta en el futuro = Erhaltung des Kulturerbes im digitalen Zeitalter : eine gemeinsame Aufgabe von Medienspezialisten*. 2002-02-15. URL: <http://www.payer.de/einzel/patrimonio.htm>. Zugriff am: 2004-05-07

Payer 2002b Payer, Margarete: *Computervermittelte Kommunikation*. 2002-05-24: URL: <http://www.payer.de/cmc/cmcs0.htm>. Zugriff am:2004-08-13

Payer 2002c Payer, Margarete: *Datenbankaufbau*. 2002-05-14. URL: <http://www.payer.de/dblink.htm>. Zugriff am: 2004-08-21

Payer 2002d Payer, Margarete: *Formate in bibliographischen Datenbanken*. In: *Datenbankaufbau : Kapitel 7*. 2002-05-14. URL: <http://www.payer.de/dbaufbau/dbauf07.html>. Zugriff am: 2004-08-21

Payer 2002e Payer, Margarete: *Die Welt neben AACR2 und RAK-WB*. 2002-07-09. URL: <http://www.payer.de/einzel/weltnebenrak.htm>. Zugriff am: 2004-08-14

Payer 2003 Payer, Margarete: 6B.8.2.2. PGP. In: *Computervermittelte Kommunikation*. 2003-10-08: URL: <http://www.payer.de/cmcs/cmcs06b.htm>. Zugriff am: 2004-08-13

RLG/OCLC 2002 Research Libraries Group und Online Computer Library Center: *Trusted Digital Repositories : Attributes and Repositories*. 2002-05. URL: <http://www.rlg.org/longterm/repositories.pdf>. Zugriff am: 2004-07-19

Rothenberg 2000 Rothenberg, Jeff: *An Experiment in Using Emulation to preserve Digital Publications*. 2000-07. URL: <http://www.kb.nl/coop/nedlib/results/emulationpreservationreport.pdf>. Zugriff am: 2004-05-01

Scheffczyk 2004 Scheffczyk, Jan: *Wie erstellen wir haltbare Dokumente?*. 2004-03-24. URL: http://www.langzeitarchivierung.de/downloads/BK_Leipzig_LZA_Scheffczyk.pdf. Zugriff am: 2004-04-10

Scholz 2004 Scholz, Dirk: *Kriterien für vertrauenswürdige Archive*. 2004-03-24. URL: http://www.langzeitarchivierung.de/downloads/Leipzig2004_Scholz.ppt. Zugriff am: 2004-04-10

Scholze 2002 Scholze, Frank: *Authentizität und Langzeitarchivierung*. In: *Workshop „Langzeitverfügbarkeit digitaler Dokumente – Erarbeitung eines ersten kooperativen Konzepts für Deutschland“ : Arbeitspapier*. 2002.10-04. URL: <http://www.dl-forum.de/Foren/Langzeitverfuegbarkeit/Arbeitspapier2.pdf>. Zugriff am: 2004-08-13

Schröder 2004a Schröder, Kathrin: *„Das EPICUR-Projekt Der Deutschen Bibliothek: URN-Vergabe, aber wie?“*. 2004-04-30. URL: <http://www.persistent-identifizier.de/download/EPICUR-Vortrag-KOBV-20040430.pdf>. Zugriff am: 2004-06-03

Schroeder 2004b Schroeder, Kathrin: *Persistent Identifiers (PI) im Kontext der Langzeitarchivierung (LZA) – State of the Art*. 2004-03-24.

http://www.langzeitarchivierung.de/downloads/BK_Leipzig_LZA_Schroeder.pdf.

Zugriff am: 2004-04-10

Schwens 1998 Schwens, Ute: *Sammlung und Bereitstellung elektronischer Dissertationen – die Praxis Der Deutschen Bibliothek*. 1998-11-24. URL:

<http://eldorado.uni-dortmund.de/8080/bib/98/workshop/schwens/schwens.pdf>.

Zugriff am: 2004-06-01

Schwens 2002 Schwens, Ute: Die Deutsche Bibliothek – gesetzlicher Auftrag und elektronische Publikationen. In: *Zeitschrift für Bibliothekswesen und Bibliographie* Jg. 49 (2002), Nr. 1, S. 13 – 17

Schwens 2004 Schwens, Ute: *NESTOR : Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen in Deutschland*. 2004-06-02. URL:

http://www.isn-oldenburg.de/nestor-workshop/slides/slides_schwens.pdf. Zugriff am: 2004-09-06

Springer-Verlag und Die Deutsche Bibliothek kooperieren in der Langzeitarchivierung von Online-Publikationen 2002 Springer-Verlag und Die Deutsche Bibliothek kooperieren in der Langzeitarchivierung von Online-Publikationen. [2002-10]. URL: http://voeb.uibk.ac.at/vm/vm55-34_beitr_pressemitteilungen.pdf.

Zugriff am: 2004-10-03

Steinke 2004 Steinke, Tobias: *Neuer Lieferweg für Hochschulschriften über OAI-Schnittstelle*. 2004-03. URL:

<http://www.dissonline.de/tagungen/Bkongress2004/Bibkongress-OAI.ppt>. Zugriff

am: 2004-08-28

Steinke und Altenhöner 2004 Steinke, Tobias ; Altenhöner, Reinhard: *Ein Depotsystem zur Langzeitarchivierung digitaler Publikationen : von der Theorie zur Praxis*. 2004-03-24. URL:

http://www.langzeitarchivierung.de/downloads/BK_Leipzig_LZA_Steinke_Altenhoener.pdf. Zugriff am: 2004-04-10

Tappenbeck 2001 Tappenbeck, Inka: *Carmen AP 2/5 : Metadaten zu Terms & Conditions und zur Archivierung*. 2001-02-01. URL: <http://www.sub.uni-goettingen.de/carmen/Veroeffentlichungen/carmenworkshop.ppt>. Zugriff am: 2004-08-23

Tiedau 2003 Tiedau, Ulrich: *Kompetenznetzwerk Langzeitarchivierung in Deutschland*. [2003]. URL: <http://voeb.uibk.ac.at/odok2003/tiedau.pdf>. Zugriff am: 2004-09-06

UNESCO 2003a United Nations Educational, Scientific and Cultural Organization: *Charter on the Preservation of the Digital Heritage*. 2003-10-17. URL: http://portal.unesco.org/ci/ev.php?URL_ID=13366&URL_DO=DO_TOPIC&URL_SECTION=201&reload=1070274443. Zugriff am: 2004-06-06

UNESCO 2003b United Nations Educational, Scientific and Cultural Organization: *Gleichberechtigter Zugang zum Wissen und Stärkung der Meinungsfreiheit*. 2003-10/11. URL: <http://www.unesco-heute.de/1103/gkcii.htm>. Zugriff am: 2004-06-05

van der Werf 2000 van der Werf, Titia: *The Deposit System for Electronic Publications*. 2000-11. URL: <http://www.kb.nl/coop/nedlib/results/DSEPprocessmodel.pdf>. (NEDLIB Report series ; 6). Zugriff am: 2004-07-28

Weiß 2004 Weiß, Berthold: *Dauerhafte Archivierung und Authentifizierung von WWW-Volltexten – Die Vorgehensweise Der Deutschen Bibliothek bei der Sammlung von Netzpublikationen*. 2001-12-05. URL: http://www.dipf.de/ueber_uns/weiss.ppt. Zugriff am: 2004-04-25

Wiesenmüller 2004 Wiesenmüller, Heidrun: Langzeitarchivierung von Online-Publikationen an Regionalbibliotheken: Das Projekt ‚Baden-Württembergisches Online-Archiv‘ (BOA). In: *Bibliotheksdienst* Jg. 38 (2004), Nr. 4, S. 471 – 479

Winkler 2004 Winkler, Peter: PC-Lexikon 2004 / Peter Winkler

München : Markt+Technik-Verl., 2004. - 1004 S.

ISBN 3-8272-6646-7

Wollschläger 2003 Wollschläger, Thomas: Aufbau einer Koordinierungsstelle für Online-Hochschulschriften. In: *Dialog mit Bibliotheken* Jg. 15 (2003), Nr. 3, S. 50 – 53

Wollschläger 2004 Wollschläger, Thomas: 3. Workshop DissOnline in Der Deutschen Bibliothek. In: *Dialog mit Bibliotheken* Jg. 16 (2004), Nr. 1, S. 55 – 56

Workshop 2002 Workshop „Langzeitverfügbarkeit digitaler Dokumente – Erarbeitung eines ersten kooperativen Konzepts für Deutschland“ : Arbeitspapier. 2002.10-04. URL: <http://www.dl-forum.de/Foren/Langzeitverfuegbarkeit/Arbeitspapier2.pdf>. Zugriff am: 2004-04-16

WU 2002a Wirtschaftsuniversität Wien: *Temporales Web-Management: Zielsetzungen der Dublin Core Metadata Initiative*. 2002-11-05. URL: <http://www.wu-wien.ac.at/usr/ebner/archive/diss/html/node37.html>. Zugriff am: 2004-08-10

WU 2002b Wirtschaftsuniversität Wien: *Temporales Web-Management: Qualifiziertes Dublin Core*. 2002-11-05. URL: <http://www.wu-wien.ac.at/usr/ebner/archive/diss/html/node39.html>. Zugriff am: 2004-08-10

Zheng 2003 Zheng, Wang: *Erstellung digitaler Dokumente*. 2003-01-15. URL: http://www.ipd.uka.de/~oosem/DB0203/folien/wang_zheng.pdf. Zugriff am: 2004-07-31

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Diplomarbeit selbständig angefertigt habe. Es wurden nur die in der Arbeit ausdrücklich genannten Quellen und Hilfsmittel benutzt. Wörtlich oder sinngemäß entnommenes Gedankengut habe ich als solches kenntlich gemacht.

Ort und Datum

Unterschrift